APR 18 1929

PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS
HAROLD RUGG, *Chairman.*

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — L. M. TERMAN
FRANK N. FREEMAN — V. A. C. HENMON — BEARDSLEY RUML — E. L. THORNDIKE

VOL. XX
No. 4

APRIL, 1929

$5.00 a year.
70c per copy.

CONTENTS

THE RELATIVE EFFECT OF NATURE AND NURTURE INFLUENCES ON TWIN DIFFERENCES. *Karl J. Holzinger* 241

CHILDREN'S INTEREST IN COLLECTING. *M. T. Whitley* 249

AN EXPERIMENTAL STUDY OF BRIGHT AND DULL HIGH SCHOOL PUPILS. *W. H. Pyle and G. H. Snadden* 262

ARE BOYS EXCELLING GIRLS IN GEOMETRIC LEARNING? *Winona M. Perry* 270

ERRONEOUS FIRST IMPRESSIONS ON OBJECTIVE TESTS. *C. O. Mathews* 280

THE INFLUENCE OF AFFECTIVE TONE ON LEARNING AND RETENTION. *Ruth M. Chaney and Alvhh R. Lauer* 287

TRANSFER OF TRAINING IN THE CANCELLATION EXPERIMENT. *B. H. Newkirk and R. H. Gundlach* 291

THE CONSTANCY OF IMAGINAL CONTENT. *Henry Bowers* 295

SOME RESULTS OF THE MULTI-MENTAL TEST. *Paul A. Witty and J. F. Taylor* 299

PERFORMANCE TESTS AND FOREIGN CHILDREN. *Olive P. Lester* 303

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES 310

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION. 316

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street — Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

HARVARD UNIVERSITY

Summer School of Arts and Sciences and of Education

JULY 8—AUGUST 17, 1929

The Psychology of Mentally Deficient, Retarded, and Psychopathic Children

Course for School Psychologists and Special Class Teachers

Lectures, clinics, demonstration classes for the study of the methods of training given by Professor Walter F. Dearborn, with the coöperation of Dr. Leonhard Seif, Neurologist and Director of Child Guidance Clinic, Munich, Germany; Dr. Ransom A. Greene, Superintendent of the Walter E. Fernald State School, Waverley, Massachusetts; Dr. Karl M. Bowman, Chief of Staff, Boston Psychopathic Hospital, and of other specialists in this field.

For information concerning this course and other courses in Education and related subjects, address the

Secretary of the Summer School
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

Educational Administration and Psychology

Including Teacher Training

is published monthly, except June, July, and August by Warwick & York, Inc., Baltimore, Md., and York, Pa. Title page and index are bound in the December number.

The price of the Journal is $4.00 a year, payable in advance. Foreign postage is 30 cents extra. Single current issues are 60 cents each and less than a full year's subscription is charged at single issue rate for every month ordered. Back volumes $4.50 each; single back issues 75 cents.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which such change is to take effect. No claim for non-receipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Manuscripts for publication may be submitted to any one of the Editors or to the publishers direct. Books for review should be addressed to the publishers, 10 E. Centre Street, Baltimore, Md.

Warwick & York, Inc.,
Baltimore, Md.

TEACHERS' HEADQUARTERS

European Plan

Entirely Fire-Proof

Centrally Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

SCHOOL AND SOCIETY

A weekly journal covering the field of education in relation to the problems of American democracy.

The objects of SCHOOL AND SOCIETY are the advancement of education as a science and the adjustment of our lower and higher schools to the needs of modern life.

It encourages discussion of all problems from every point of view; it pays special attention to foreign affairs; it publishes accounts of educational events of the past week and research on fundamental problems.

Subscription $5.00—Single Numbers 15 Cents

[*SCHOOL AND SOCIETY will be sent to new subscribers for thirteen weeks for one dollar, which may be sent at the risk of the publisher.*]

THE SCIENCE PRESS
Grand Central Terminal - - New York, N. Y.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XX | April, 1929 | Number 4

THE RELATIVE EFFECT OF NATURE AND NURTURE INFLUENCES ON TWIN DIFFERENCES

KARL J. HOLZINGER

University of Chicago

The Data

The data employed in this study consist of a portion of the material gathered for an extensive investigation of identical and fraternal twins. The measurements were made by Professor H. H. Newman (Department of Zoology), Professer F. N. Freeman, Mrs. Bryan Mitchell, and the writer, all of the University of Chicago. A number of articles and monographs based on these data are now in preparation.

Identical (monozygotic) twins are those resulting from the fertilization of one egg by a single sperm and from such origin may be considered as having the same heredity. Fraternal (dizygotic) twins are those from two ova fertilized by separate sperms—the embryos developing side by side, each with its own placenta and chorin. For the latter type of twins the heredities will be different. Professor Newman, who is an eminent authority on the biology of twinning, was chiefly interested in the separation of twins into these two classes. A large number of criteria were used by him in classifying twins according to these two types, a full report of which will appear later. The final number of identical twin pairs secured was fifty, while the number of fraternal (like-sexed) twin pairs obtained was fifty-two. These numbers, of course, do not correspond to the proportions of the two types in the general population, but the data were increased to get samples of comparable size.

The tests made consisted of a large number of physical and mental measurements, palm prints, records of physical likeness and peculiarities, scholastic standing, and considerable data on interests in social and mental activities. The chief measurements used in the present

TABLE I.—SUMMARY OF CERTAIN CONSTANTS

Constant	Identical, 50 pairs	Fraternal, 52 pairs	Difference
Mean age	13 yr., 4.1 mo.	13 yr., 5.4 mo.	−1.3 ± 3.3
SD age	34.9 mo.	34.0 mo.	.9 ± 2.3
Mean school grade	7.9	7.9	0.0
Mean finger ridges (both)	83.2	74.0	9.2 ± 2.4
Mean cephalic index	.795	.800	− .005 ± .004
Mean standing height	151.6 cm.	152.1 cm.	− .5 ± 1.4
Mean weight	97.9 lb.	98.6 lb.	− .7 ± 2.6
Mean Binet IQ	101.6	96.1	5.5 ± 1.6
SD Binet IQ	17.3	15.7	1.6 ± 1.1
Mean Otis IQ	103.7	99.0	4.7 ± 1.5
Mean EA	13 yr., 4.4 mo.	13 yr., 2.0 mo.	2.4 ± 3.0
Mean motor	308.1	307.2	.9 ± 4.5
Mean Woodworth-Mathews	14.7	14.9	− .2 ± .8

TABLE II.—CORRELATIONS (AGE CONSTANT)

Variable	Identical N = 50 (pairs)	Fraternal N = 52 (pairs)
Standing height[1]	.93 ± .01	.65 ± .05
Sitting height	.88 ± .02	.50 ± .07
Weight	.92 ± .01	.63 ± .06
Head length	.91 ± .02	.58 ± .06
Head width	.89 ± .02	.54 ± .07
Cephalic index[1]	.89 ± .02	.58 ± .06
Finger ridges (both)[1]	.97 ± .01	.46 ± .08
Binet MA[1]	.86 ± .02	.60 ± .06
Binet IQ	.88 ± .02	.63 ± .06
Otis score	.90 ± .02	.57 ± .06
Otis IQ	.92 ± .01	.62 ± .06
Word meaning	.86 ± .02	.56 ± .06
Arithmetic	.73 ± .04	.69 ± .05
Nature study	.77 ± .04	.65 ± .05
History and literature	.82 ± .03	.67 ± .05
Spelling	.87 ± .02	.73 ± .04
All, EA[1]	.89 ± .02	.70 ± .05
Woodworth-Mathews[1]	.56 ± .07	.37 ± .08
Downey judgment	.31 ± .08	.37 ± .08
Downey motor index	.51 ± .07	.53 ± .07
Downey coordinated impression	.48 ± .07	.51 ± .07
Tapping total[1]	.69 ± .05	.38 ± .08

[1] Values are variables studied in detail later.

study are indicated in Table I and II. These include (1) total number of finger ridges (from the tri-radius) on both hands, (2) cephalic index, (3) standing height, (4) mental age on the Stanford-Binet individual test, (5) educational age on the Stanford Achievement Tests, (6) motor test giving rates of tapping with finger and wrist (results from both hands combined), and (7) the Woodworth-Mathews Questionnaire which is a measure of emotional stability.

General Characteristics of the Two Groups

The identical and fraternal groups appear to be very much alike in certain general characteristics indicated in Table I. The agreement in mean age, mean grade, mean height and weight is remarkable. From the differences and their probable errors given in the last column it is apparent that the intelligence and total number of finger ridges of the mean values for the identical group are significantly larger than those for the fraternal group. Professor Terman found for 905 "unselected" children that the mean IQ was 101.5 ± 0.3 with a standard deviation of 13.0 ± 0.2. If the corresponding constants for the identical twins be compared with these values we find

$$\text{Identical mean IQ} - \text{Terman mean IQ} = 0.1 \pm 1.2$$
$$\text{Identical SD} - \text{Terman SD} = 4.3 \pm 0.9$$

We may therefore conclude that identical twins here studied have the same average intelligence as unselected children, but are somewhat more variable.

The correlations in Table II give evidence of the closer genetic relationship of identical as compared with fraternal twins. The typical correlation (holding age constant by the usual formula) is about .90 while the average correlation for fraternal twins is approximately .60. Some of the coefficients are smaller than these values partly on account of the unreliability of certain of the tests. For the more reliable of the mental tests such as Binet and the Stanford Achievement Test, the correlations are approximately equal to those found in the case of physical traits such as height and weight.

The Relative Importance of Nature and Nurture in Determining Mean Differences

The chief purpose of this paper is to present certain formulas for the study of the relative effect of nature and nurture upon twin

differences. These formulas will then be applied to the twin data and the relative importance of the two factors determined for certain physical and mental traits.

It is first assumed the heredity of the twins classified as *identical* is sufficiently similar so that a difference between a pair of twins in a certain trait may be thought of as almost entirely due to nurture designated as B. A difference between pair of fraternal twins for this trait is considered as a linear function of B and of a nature factor A. Identical and fraternal twin differences may then be set down as follows:

Identical	Fraternal
$_iX = {}_iA + {}_iB + {}_iE$	$_fX = {}_fA + {}_fB + {}_fE$
$_iX' = {}_iA' + {}_iB' + {}_iE'$	$_fX' = {}_fA' + {}_fB' + {}_fE'$
$_iD = {}_iB' - {}_iB = {}_iD_b$	$_fD = {}_fD_a + {}_fD_b$

The front subscripts i and f refer to identical and fraternal twins. Primes are used to distinguish members of a pair, E is an accidental error (to be subsequently ignored). D is difference in the trait. D_a is a difference due to nature and D_b is a difference due to nurture.

The first formula may then be written

$$T = \frac{{}_f\bar{D} - {}_i\bar{D}}{{}_i\bar{D}} = \frac{{}_f\bar{D}_a + {}_f\bar{D}_b - {}_i\bar{D}_b}{{}_i\bar{D}_b} = \frac{{}_f\bar{D}_a}{{}_f\bar{D}_b} \qquad (1)$$

where $\bar{D}$ is the mean absolute difference, $X - X'$. It has been assumed that the mean nurture difference $\bar{D}_b$ is the same for identical and fraternal twins. Formula (1) gives an approximate measure of the relative effectiveness of nature and nurture in determining mean twin differences.

It will be noted that nature, or A, is a factor here considered as the same for identical twins,[1] but different for fraternal twins. The factor A may, of course, change with age and differ for various twin pairs, but it is assumed that for a given identical twin pair it is the same. The B factor, on the other hand, represents the effect of environment or nurture. Such environment is considered as the sum of all influences producing differences in twins effective

[1] If the separation of the twins into the identical and fraternal classes is not exact, or if identical twins have only approximately the same heredity, ${}_iD$ will be a function of ${}_iD_b$ and a small value ${}_iD_a$. The presence of this additional ${}_iD_a$ quantity will make obtained values of T slightly smaller than the true values.

before and after birth. As an example, take the intelligence of a pair of identical twins. With only nature, or A, operative, these two would be identical in intelligence at any age, but with the additional influence of nurture, or B, a difference between the two will be revealed by any measurement, and such a difference is here regarded as a result of differential effect of B upon each twin. A value of 1.00 from formula (1) indicates that nature and nurture have been equally effective in producing mean twin differences.

A few mean differences have been set forth in Table III for illustration. The differences are much greater in the case of fraternal twins and there is a little evidence that differences are greater for older than for younger twins. The values of T from formula (1) are given in Table IV. For finger ridges and height nature is more important than nurture, while for the four mental traits nurture appears slightly more effective than nature. There is also some evidence that nurture is more effective for younger than for older twins. The above results are not obtained from large samples and the probable

TABLE III.—MEAN DIFFERENCES BETWEEN PAIR MEMBERS

Variable	Identical			Fraternal		
	Younger	Older	All	Younger	Older	All
Finger ridges (both)	3.2	3.6	3.4	21.7	22.9	22.4
Cephalic index	...	...	.02	....	....	.03
Height	1.7	1.6	1.7	4.4	4.4	4.4
MA	7.2	9.8	8.4	10.6	19.9	15.9
EA	5.5	7.5	6.5	6.5	15.1	11.6
Motor	...	...	19.0	....	....	29.0
Woodworth-Mathews	...	...	5.6	....	....	6.8

TABLE IV.—VALUES OF T FROM FORMULA (1)

Variable	Younger	Older	All
Finger ridges (both)	5.8	5.4	5.6
Cephalic index	...	...	.7
Height	1.6	1.7	1.7
MA	.5	1.0	.9
EA	.2	1.0	.8
Motor	...	...	.5
Woodworth-Mathews	...	...	.2

errors have not been found, but the consistency of the findings is very suggestive.[1]

The Relative Importance of Nature and Nurture in Determining Variability of Twin Differences

Instead of studing the mean twin differences of a group, we may investigate the variability of these differences. More particularly we shall derive a formula which gives approximately the ratio of the standard deviation of D_a to the standard deviation of D_b.

Let

r = the correlation between X and X',
R = the correlation between D_a and D_b,
S = the standard deviation of X,
S_d = the standard deviation of D,
S_a and S_b = the standard deviations of D_a and D_b, all values being found from double entry tables.

We may then write

$$_iS_d^2 = {}_iS_b^2 \quad \text{and} \quad {}_fS_d^2 = {}_fS_a^2 + {}_fS_b^2 + 2R\,{}_fS_a \cdot {}_fS_b$$

Two familiar formulas relating variance to correlation may be readily derived and written in the form

$$\frac{{}_iS_d^2}{{}_iS^2} = 2(1 - {}_ir) \tag{2}$$

$$\frac{{}_fS_d^2}{{}_fS^2} = 2(1 - {}_fr) \tag{3}$$

We next set up a function, t^2, involving the variance of D relative to the variance of X. Thus

$$t^2 = \frac{\dfrac{{}_fS_d^2}{{}_fS^2} - \dfrac{{}_iS_d^2}{{}_iS^2}}{\dfrac{{}_iS_d^2}{{}_iS^2}} = \frac{{}_fS_a^2 + 2R\,{}_fS_a \cdot {}_fS_b}{{}_fS_b^2} \tag{4}$$

If the correlation R between D_a and D_b is zero, this reduces to

$$t^2 = \frac{{}_fS_a^2}{{}_fS_b^2} = U^2 \tag{5}$$

If R is not zero, we may write

$$t^2 = U^2 + 2RU \tag{6}$$

[1] A much more detailed report will appear in a forthcoming monograph.

A still more convenient form of formula (4) may be found by substituting therein the values given by formulas (2) and (3). This yields

$$t^2 = \frac{(1 - {}_fr) - (1 - {}_ir)}{(1 - {}_ir)} = \frac{{}_ir - {}_fr}{1 - {}_ir} \qquad (7)$$

This last formula is very easy to apply and is capable of simple interpretation. If no nurture factor, B, were operative in the case

TABLE V.—CORRELATIONS AND VALUES OF t^2 FROM FORMULA (7)
(All Coefficients Corrected for Age)

Variable	Correlation		t^2	PE of t^2
	Identical	Fraternal		
Standing height	.93	.65	4.0	1.2
Weight	.92	.63	3.6	1.1
Head length	.91	.58	3.7	1.1
Total finger ridges	.97	.46	17.0	4.2
Binet MA	.86	.60	1.9	0.7
Binet IQ	.88	.63	2.0	0.7
Word meaning	.86	.56	2.1	0.7
Arithmetic	.73	.69	0.2	0.3
Nature study	.77	.65	0.5	0.3
History and literature	.82	.67	0.8	0.4
Spelling	.87	.73	1.1	0.5
EA	.89	.70	1.7	0.7
Woodworth-Mathews	.56	.37	0.4	0.3
Motor	.69	.38	1.0	0.4

of identical twin differences, then the correlation ${}_ir$ would be +1.00. The quantity $1 - {}_ir$ therefore measures the effect of the B factor alone. In the case of fraternal twins, both A and B factors are operating and $1 - {}_fr$ measures the combined influence of the two. Formula (7) therefore measures the relative influence of A to B factors in producing differentiation.

The probable error of t^2 from formula (7) has just been derived and is given here without proof.

$$\text{PE of } t^2 = .6745\frac{1 - {}_fr}{1 - {}_ir}\sqrt{\frac{(1 + {}_ir)^2 + (1 + {}_fr)^2}{N}} \qquad (8)$$

Illustrations of the use of formula (7) are given in Table V. It will be noted that nature is much more effective than nurture for physical traits. In the case of mental age, IQ, and word-meaning,

the value of t^2 is approximately 2.00 with a disconcerting large probable error. There is some evidence from the regularity of these values, however, that nature is somewhat more effective than nurture in the determination of difference variability for mental traits.

In the case of achievement tests, the values of t^2 are not very significant, but their trend indicates that nature is hardly more effective than nurture in determining difference variability.

Summary

Formulas for studying the relative effect of nature and nurture upon mean twin differences and their variability.

The application of these formulas to the present twin data suggests the following conclusions:

1. For intelligence, nature is about equally effective with nurture in producing mean twin differences, but nature is somewhat more effective than nurture in determining variability of such differences. Nurture influences would thus seem to play a large part in the modification of intellect.

2. In the case of scholastic achievement, the relative influences of nature and nurture probably changes with the trait. On the whole, nature would appear little, if any, more effective than nurture. The question has often been asked, "What per cent of a person's success in achievement is due to nature and what per cent is due to nurture?" Such a percentage comparison may be impossible and even meaningless, but the above results would suggest a 50–50 division rather than a 97–3 division as given by Courtis in his book, "Why Children Succeed."

CHILDREN'S INTEREST IN COLLECTING

M. T. WHITLEY

Teachers College, Columbia University

As a result of some findings from their play quiz[1] Lehman and Witty have stated their conclusion that children nowadays are less interested in making collections than they were at the time Burk's investigations were made, thirty years ago.[2] They further suggest that collecting was somewhat of the nature of a passing fad, brought into prominence and unduly emphasized in its educational significance by the mere fact that Burk's questionnaire synchronized with the crest of the wave of interest in that fad. They parallel this with the conclusions that might have been drawn by an investigation in 1923–1924 into children's interest in crossword puzzles.

Granted that these assumptions are correct, the advice to make educational use of such interest, so often met with in books on pedagogy, is doubtless out of date—no longer valid if the fad has died out. In the mind of the writer, however, the assumptions are not perfectly sound. To begin with, the authors' only basis—at least, the only one published, whatever other details their immense quantities of first-hand data may contain—is the "yes" or "no" answers in the form of a check in front of item forty-one, "collecting stamps, birds' eggs and so on."

This is one of two hundred questions, or items, on play interests as a whole. The general directions for checking read as follows:

"What things have you been doing during the past week just because you wanted to? Read through the following list and . . . draw a circle . . . around each number that stands in front of anything . . . that you have been doing during the past week just because you wanted to do it."

Now the idea occurs that this particular form would not bring many affirmative answers for four reasons. First, the months in which the play quiz was used were November, January, February, and April. For children in and around Kansas City, it would be a seasonal improbability to collect many birds' eggs in the weeks preceding those dates. Second, the interest in stamp collecting did not, according to

---

[1] Lehman, H. C., and Witty, P. A.: "The Psychology of Play Activities."

[2] Burk, C. F.: The Collecting Instinct. *Pedagogical Seminar*, 1900.

Burk's returns, hit the majority of children, and then chiefly between the ages of nine and fourteen.

Third, in the writer's experience, the term "and so on" is frequently ignored even by adults, let alone children in a hurry to proceed with other interesting items dealing with games. One wonders what results on play activities during a week in winter would have been secured by Lehman and Witty if their sole question to all ages and both sexes among ninety-nine others on all sorts of subjects, had been "Have you, during the last week, played house or gone swimming?" Fourth, in class reports by students of simple investigations on collecting an unexpected difficulty had been found in children's non-understanding of the word "collection." It seemed to connote a formal effect, arranged, cataloged and displayed, rather than to include their informal hoards and casual impulses to possess themselves of trifles. In oral work the need was found to interpret by the phrase "Have you any little keepsakes?" when a puzzled "no" to the earlier inquiry was often replaced by an eager "yes."

Such oral reports had been customary for several years merely for the interest of the students in the child psychology classes at Teachers College; but in the winter term of 1927 it was decided to undertake something yielding more definite returns. Accordingly, a questionnaire covering sixty items was drawn up, modeled frankly on the play quiz, but dealing solely with kinds of things collected. The period was changed to cover two weeks, as it was believed that many children during school weeks do not have constant opportunities to add to their collections, certainly not as frequent chances as to play games of their liking. The dates on which it was used came about ten days after the Thanksgiving vacation which further helped to set off the period in the minds of the children. The questionnaire was given by students in the Saturday section of the course in child psychology at Teachers College who were at the time engaged either in classroom teaching or as school principals. Another class gave assistance in tabulating and percentaging the returns. A few interested parents and aunts submitted a small number of returns, mainly from younger children. In some cases this was used as a class exercise. In no case were the children hurried. For Grade III and below, in individual work, roughly wherever advisable, the questions were read aloud; indeed, every assistance was given to insure a thorough understanding of the requirements.

An abbreviated form is shown here as a sample.

Name.....Age.....Grade.....
School.....City.....State.....

I. Have you any treasures, keepsakes, things you are collecting?.....

II. Read through the list below. If you have, *during the last two weeks*, collected any of the things there, put a check (✓) by it.

1. pieces of string	21. funny papers	41. stones, pebbles, rocks
2. pieces of cloth	22. coupons	42. seeds
3. rubber bands	23. stamps	43. butterflies, moths
4. pieces of metal	24. cigar bands	44. insects, bugs
5. small boxes	25. tobacco labels	45. shells
	etc.	

Add anything, not on the list, which you like to keep, and have collected during the past two weeks.

....
....

III. What do you do with the things you collect?.....

keep to look at sometimes	give away
use in play	trade with others
show to friends	use in school work
keep in case you'll need them	nothing particular

IV. Where do you keep them?.....

box	drawer	cupboard
shelf	book	your own room

Have you a "memory book" for favors, invitations, place cards, programs, etc., that remind you of good times you have had?.....

V. How do you get these things you like to keep?.....

hunt for them	have them given you
buy them	trade with other people
just take them	pick them up when you find them

Have you ever written to any advertisers for free samples?.....

VI. When do you generally add to your collection?.....

any time at all	on vacation
on holidays	in the summer
special season of the year	whenever the chance comes

VII. If not now collecting any of the things in the list, did you ever do so?
At what age?.....
Which? (Enter the numbers from the list).....

Answers were secured from children in thirty-four different towns and cities in the states of Pennsylvania, Connecticut, New Jersey, New York and West Virginia. The largest number from any one school was four hundred forty-eight, the smallest ten.

A total of 4446 returns was considered, some dozen or more having been discarded when over-age children, presumably of low IQ had cheerfully checked from fifty-five to sixty items on the list.

In detail, we had returns by ages as follows:

TABLE I

Age	Boys	Girls	Total	Age	Boys	Girls	Total
7 or under	69	34	103	14	203	198	401
8	106	117	223	15	163	139	302
9	195	208	403	16	108	83	191
10	251	300	551	17	42	117	159
11	321	343	664	18 or over	34	230	264
12	333	318	651	Totals.....	2097	2349	4446
13	272	262	534				

This is almost four times as many children as were included in Burk's investigation, though nothing like so large a number as were used in the play quiz.

The next table shows the median number of items checked at each age by both boys and girls, also the Q_1, the Q_3 and the maximum for both sexes together. The minimum, of course, is zero.

TABLE II

Age	Boys, Md	Girls, Md	Both sexes			
			Q_1	Md	Q_3	Max
7 or less	3	2	1	2	10	31
8	6	6	2	6	11	44
9	7	8	3	7	15	40
10	9	8	4	9	18	43
11	9	8	4	9	17	44
12	9	9	4	9	12	41
13	8	8	4	8	14	45
14	7	7	3	7	13	40
15	6	7	3	6	12	36
16	5	6	3	5	9	32
17	5	5	3	5	9	28
18 plus	5	7	4	7	10	27

From the foregoing it scarcely seems as though the interest in picking up and hoarding things was missing in the present generation of school children, at all events in the places mentioned.

The next table shows, for each age, the five items most frequently checked, boys and girls separately.

TABLE III

Seven and under		Eight years	
Boys	Girls	Boys	Girls
Marbles	Funny papers	Marbles	Paper dolls
Coupons	Samples of school work	Samples of school work	Samples of school work
Old magazines	Paper dolls	Coupons	Coupons
Small boxes	Small boxes	Old magazines	Small pictures
Buttons	Rubber bands	Funny papers	Pieces of ribbon
Nine years		**Ten years**	
Marbles	Samples of school work	Coupons	Paper dolls
Coupons	Beads	Marbles	Coupons
Funny papers	Paper dolls	Stamps	Letters received
Coins	Funny papers	Rubber bands	Beads
Stamps	Coupons	Coins	Small pictures
Eleven years		**Twelve years**	
Coupons	Coupons	Coupons	Letters received
Marbles	Letters received	Marbles	Coins
Stamps	Old magazines	Coins	Samples of school work
Funny papers	Beads	Stamps	Old magazines
Rubber bands	Samples of school work	Rubber bands	Coupons
Thirteen years		**Fourteen years**	
Coupons	Letters received	Coins	Letters received
Coins	Beads	Coupons	Photographs
Marbles	Funny papers	Marbles	Samples of school work
Rubber bands	Old magazines	Stamps	Beads
Stamps	Small pictures	Samples of school work	Pictures

TABLE III.—*Continued*

Fifteen years		Sixteen years	
Boys	Girls	Boys	Girls
Coupons	Letters received	Coupons	Letters received
Coins	Photographs	Samples of school work	Photographs
Stamps	Old magazines	Photographs	Samples of school work
Samples of school work	Beads	Coins	Magazines
Photographs	Pressedflowers	Letters received	Small pictures
Seventeen years		**Eighteen years**	
Letters received	Letters received	Theater programs	Letters received
Photographs	Small pictures	Letters received	Old magazines
Theatre programs	Old magazines	Photographs	Samples of school work
Coins	Photographs	Badges	Small pictures
Old magazines	Coins	Old magazines	Photographs

The next table shows, for each age, the number of additional objects mentioned by each sex, beyond the sixty suggested in the list.

TABLE IV

Age	Boys	Girls	Age	Boys	Girls
7 or less	6	3	13	99	63
8	23	15	14	60	42
9	36	28	15	42	22
10	70	62	16	36	18
11	98	75	17	26	14
12	114	73	18	15	11

This strikingly illustrates the increase of interest in just acquiring and amassing things, confirming Burk's findings that the high peak of interest in collecting comes between ten and twelve. However, it is in absolute disagreement with her conclusion that girls are more varied in their tastes and that boys concentrate more. Here, at every age, more different additional items were mentioned by the boys.

Burk had a list of nearly three hundred articles mentioned. Our returns gave a total, including the sixty on the check list, of three

hundred seventy-eight articles, after classifying many together, such as separate articles of clothing. Of these extra three hundred eighteen, forty-eight were mentioned once only. Forty-three different items, such as wheel tires, radio parts, screws, tools, lumber, etc. were given by one hundred sixty-four boys and only seven girls. Forty different nature items were mentioned, distributed about evenly between boys and girls. Twenty things, to do with hunting, fighting and war were given by boys exclusively. Books and booklets were given by roughly four per cent of each sex; pens or pencils by four per cent of the girls, two per cent of the boys, doll accessories by three per cent of the girls of thirteen and under. For no other single item were the percentages significant.

Turning to the sixty items in the check list, we can trace for each age the percentage reporting on various sorts of things. First, those who said they had collected nothing at all in the past two weeks. For all ages together and both sexes, 5.3 per cent. Reversing the figures, it would seem that *at no age were fewer than eighty-four per cent uninterested in some sort or other of collecting activity.* Comparing this with the facts shown in Table II it may be concluded that though the number of collections per person diminishes with age, the number of people making collections of some kind increases from nine to eighteen years.

TABLE V

Age	7	8	9	10	11	12	13	14	15	16	17	18
0	10	15.6	16	9	7	9	6.8	9	7.6	5	2	1.7

Interpreting Tables II, IV and V together, the largest number of collections per person come between the ages nine and thirteen, while the greatest variety of interest is shown approximately at the same time.

Table takes note of only those items where the percentage is as high as fifteen for three consecutive years for either sex. Items 46, 47, 48 and 49 are combined, for parts of animals. This leaves forty-seven of the original sixty on the list; and among these we do not find birds' eggs, nor insects, etc.

Considering the item stamps, No. 23, Fig. 1[1] shows the change with age, for each sex. This is similar to the graph shown by Burk for the same thing, but quite different from that of Lehman and Witty.

[1] In all the figures given in the article, the profile for the collection interests of boys are shown by unbroken lines while that for girls is shown in a broken line.

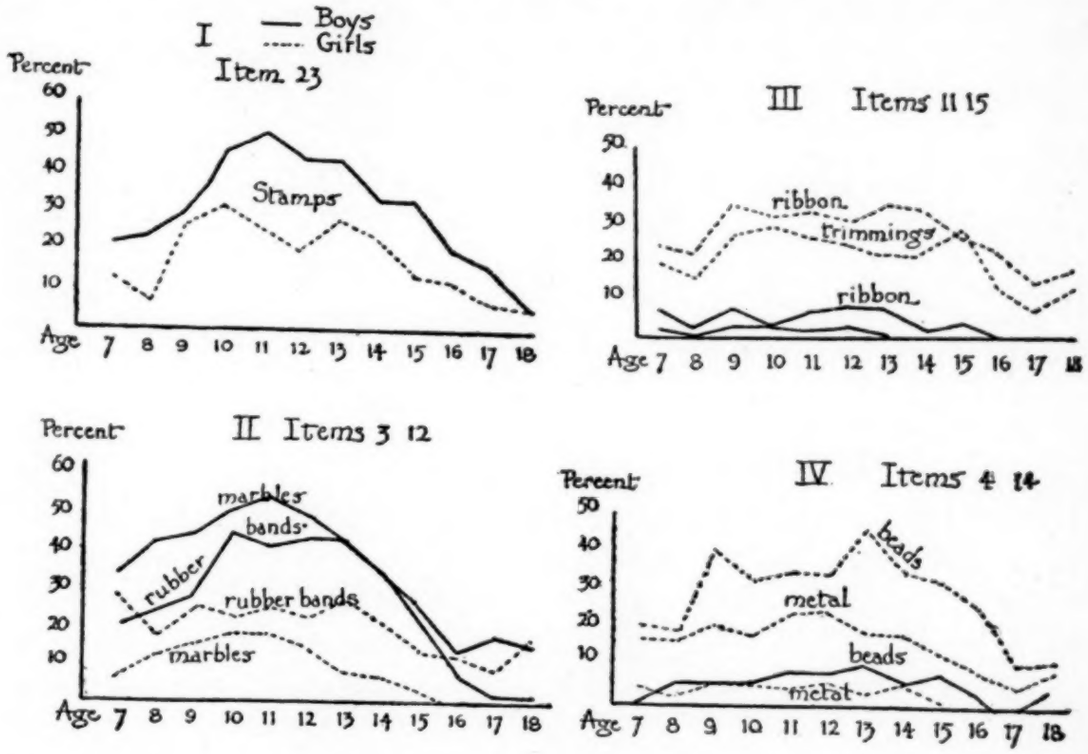
TABLE VI

	Boys												Girls											
Age	7	8	9	10	11	12	13	14	15	16	17	18	7	8	9	10	11	12	13	14	15	16	17	18
Items																								
1. Pieces of string	8	22	24	22	22	27	21	16	13	3	12	..	19	13	20	17	17	16	16	18	6	6	8	7
2. Pieces of cloth	5	5	13	15	12	12	8	5	4	5	5	6	13	19	32	30	32	24	19	21	26	6	9	12
3. Rubber bands	20	26	29	45	42	43	42	33	25	13	19	15	29	19	26	23	24	22	27	21	13	14	9	18
4. Pieces of metal	18	15	21	18	24	25	19	19	14	10	5	9	6	3	7	7	6	6	4	6	2	2	..	..
5. Small boxes	25	25	29	30	27	27	21	17	16	5	2	3	29	21	32	30	27	31	25	16	14	9	11	22
6. Tin foil	12	23	26	30	29	24	19	16	16	3	2	..	19	13	23	22	21	13	10	7	6	4	5	5
8. Buttons	25	22	23	28	24	25	30	22	13	6	14	6	16	21	26	24	24	26	25	21	14	10	8	10
9. Little ornaments	7	11	12	12	16	20	13	8	12	4	14	17	19	16	18	25	22	26	26	19	21	23	10	13
10. Little toys	17	20	27	27	26	24	12	7	7	..	2	6	26	22	33	32	27	25	16	12	6	4	2	..
11. Pieces of ribbon	7	3	8	4	7	8	8	1	3	2	..	8	26	22	37	33	34	31	35	34	26	22	14	18
12. Marbles	35	43	46	51	55	50	43	36	23	6	2	3	9	14	16	19	19	15	8	7	1	2	2	2
13. Small pictures	18	30	28	32	34	36	29	27	23	11	19	15	22	23	35	36	35	39	45	37	29	36	42	49
14. Beads	5	6	7	7	10	10	11	7	8	1	..	3	22	20	44	36	38	37	48	37	34	27	11	12
15. Pieces of trimmings	2	1	5	5	4	4	3	1	1	..	..	..	19	14	27	30	28	25	22	21	28	12	8	13
16. Autographs	2	3	9	11	12	11	11	11	13	17	17	12	..	7	6	18	20	22	22	20	25	24	14	18
17. Puzzle questions	..	9	12	13	19	20	17	13	14	7	12	3	6	7	8	13	12	12	10	11	8	7	6	9
18. Samples school work	5	36	27	32	25	36	30	34	32	35	28	29	32	29	45	35	32	40	37	42	28	43	34	53
20. Photographs	3	9	18	20	31	26	26	30	22	31	38	32	6	8	14	24	26	34	39	44	59	56	41	48
21. Funny papers	20	31	35	39	45	43	34	30	21	6	23	3	35	21	38	29	32	31	30	29	16	9	9	3
22. Coupons	29	35	44	54	58	55	55	48	48	47	26	20	16	25	37	39	41	39	43	35	28	17	11	17
23. Stamps	23	24	31	47	51	45	45	35	34	21	17	9	13	9	28	31	28	22	29	23	15	12	8	8
27. Doll furniture	..	3	3	3	2	4	2	..	..	1	..	..	19	16	22	22	15	16	9	7	3	2	2	2
28. Drawings, sketches	10	16	25	23	29	31	22	25	19	14	14	6	16	14	22	31	30	29	30	27	29	19	18	25
29. Old magazines	29	31	28	34	30	39	35	29	31	26	33	29	26	20	31	32	38	39	41	34	35	42	49	62
31. Theater programs	7	15	23	21	29	26	22	20	19	13	36	47	9	6	13	21	20	24	26	28	31	27	29	39
33. Pennants	..	3	5	9	13	14	16	14	14	13	31	18	..	4	7	5	7	7	4	6	6	7	10	12
34. Athletic trophies	1	2	5	7	14	9	12	10	14	23	21	26	..	..	2	3	4	4	3	8	7	7	5	10
35. Badges	12	23	21	19	26	24	20	14	18	9	9	3	3	5	10	15	13	14	8	7	6	5	2	3
36. Score cards	5	6	13	15	17	19	17	15	13	8	16	30	..	2	7	14	8	10	11	6	4	13	11	19
37. Letters received	12	16	24	29	28	29	30	30	28	30	43	38	13	21	32	38	40	53	55	56	69	65	67	74
38. Coins	17	26	34	39	44	48	47	52	39	30	33	26	6	19	28	34	37	40	32	29	26	14	6	11
39. Poems	5	16	19	21	20	20	16	14	13	10	19	9	10	19	25	29	26	25	30	21	26	31	41	39
40. Prizes from packages	3	19	23	24	23	18	16	12	7	4	2	9	6	15	20	23	21	20	14	16	14	9	3	4
41. Stones, pebbles, rocks	..	22	22	26	21	27	25	17	21	5	5	9	16	12	21	23	17	16	14	14	9	2	4	4
45. Shells	18	12	20	14	15	15	12	9	16	2	9	..	16	11	18	21	18	16	14	11	10	8	2	4
46 to 49. Animal parts	5	25	20	20	36	37	25	23	27	3	23	17	3	24	19	24	16	21	17	10	8	2	2	2
51. Leaves	3	10	10	14	15	19	15	11	13	6	2	6	10	13	17	22	20	23	16	17	10	1	6	5
52. Pressed flowers	5	9	9	10	12	12	9	6	9	4	9	3	10	11	15	25	17	27	27	29	31	25	14	13
53. Bottle tops	15	19	22	23	20	17	14	8	7	..	..	..	16	7	12	12	9	8	7	3	..	..	..	..
54. Spools	17	24	23	26	27	17	15	11	5	..	..	..	26	17	23	19	15	14	12	10	2	3	2	3
55. Paper dolls	2	6	6	4	2	3	2	1	1	..	..	..	26	33	44	42	36	27	19	13	5	2	8	7
56. Lists of names	2	16	16	15	15	17	16	11	15	12	19	20	3	12	23	24	26	23	26	23	25	32	13	18
59. Keys	20	25	23	36	33	32	29	29	29	16	14	9	3	10	15	14	12	16	12	10	4	5	..	..
60. Party favors	17	15	12	11	13	11	6	4	8	6	26	20	22	20	21	31	26	26	22	20	28	36	26	24

Burk further states that for all ages, sixty per cent of boys, thirty-nine per cent of girls were interested in stamp collecting. Our findings show, for all ages, thirty-eight per cent of boys, twenty two per cent of girls similarly reporting. Other comparisons with Burk are as follows:

TABLE VII.—FOR ALL AGES

	Boys		Girls	
	Burk	This study	Burk	This study
Stamps	60	38	39	22
Marbles	36	41	22	10
Buttons	14	21	23	20
Pictures, postcards, photographs	18	54	55	71
Stones, rocks, etc	11	20	9	14
Pieces of cloth	1	8	20	23

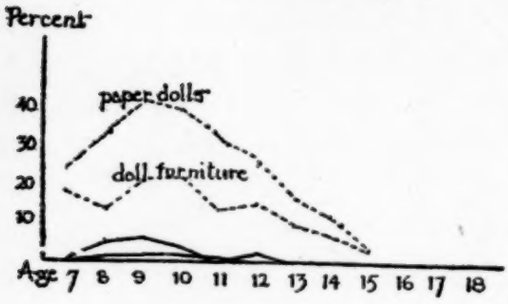


Returning to Table VI: The sentimental interest of girls may be traced by looking at No. 37, letters received. Divergent interests of boys and girls can be seen by tracing Nos. 3, 4, 11, 12, 14, 15, 27, 55. The doll interest of girls is reflected in the last two mentioned, where the age change is quite evident. More striking is the age change for item 60 for boys. Party favors evidently take on a new significance after sixteen years old.

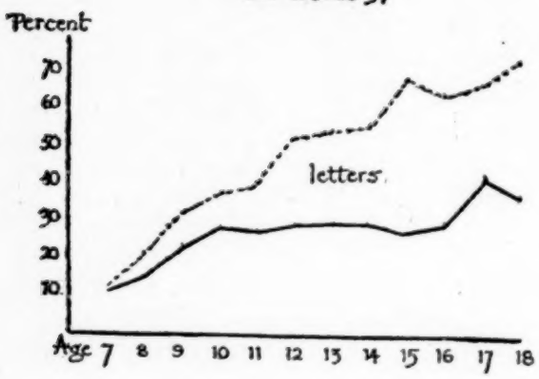
A few graphs may serve to emphasize these facts.

Lehman and Witty found that the activity "looking at the Sunday funny paper" was most frequently reported by both boys and girls

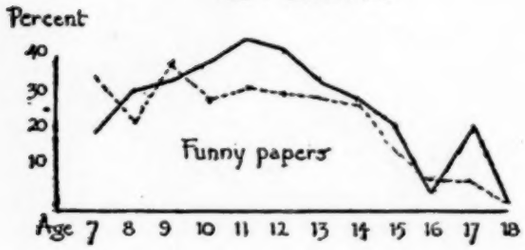
V Items 27 55



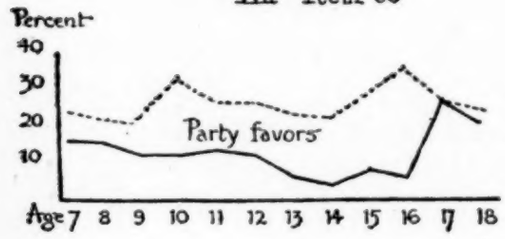
VI Item 37



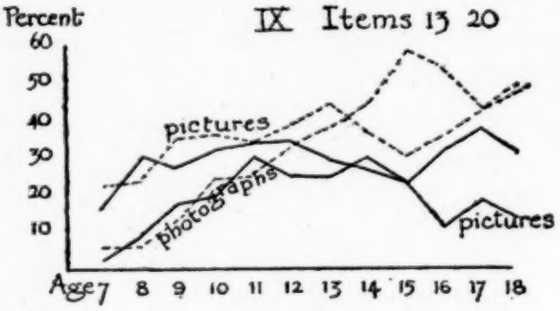
VII Item 21



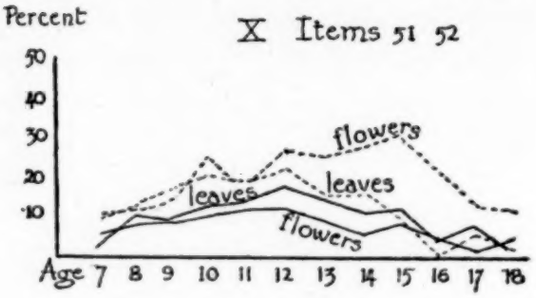
VIII Item 60



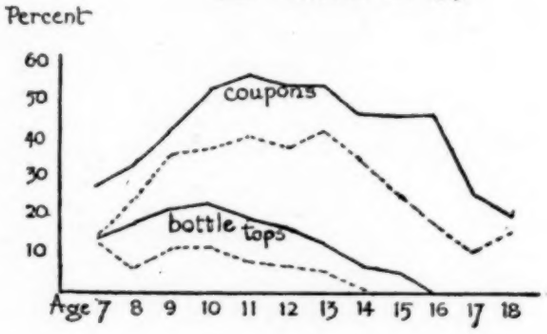
IX Items 13 20



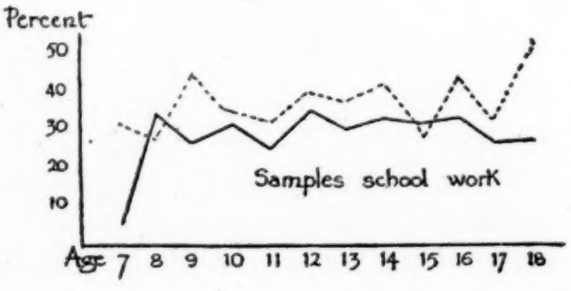
X Items 51 52



XI Items 22 53



XII Item 18



between eight and fifteen years of age. In our Table III, collecting funny papers appears among the six things most frequently mentioned at seven by girls, at eight by boys, at nine by both, at eleven by boys, thirteen by girls. Or, to trace item 21 in Table VI, it can be seen that roughly from thirty to forty per cent of children are collecting funny papers at those ages. Other comparisons of interest are item 12 in Table VI, with its high peak at eleven years old, whereas Lehman and Witty's graphs for various grouping of boys show more *playing* marbles from 8½ to 10½.

Other sex differences are shown in the answers to question V. The percentage of girls depending on gifts to add to their collections is higher at all ages than it is for boys, and consistently lower for trading. Condensed facts are given for all ages in Table VIII.

TABLE VIII

	Boys, Per Cent	Girls, Per Cent
Hunt for	34	30
Buy	33	32
Take	6	5
Gift	31	46
Trade	15	10
Pick up	37	32
Write for	40	50

The totals add to more than one hundred since most children checked several ways in which they added to their collections.

The returns to question IV (where do you keep them?) revealed little of special interest, though it acted as a rough check on the credibility of the itemized list itself. Additional places mentioned were, "in the attic," "the cellar," "an old trunk," "my desk," "a locker," "the garage," "the dog house," "under a board," "hole in the wall," "where the other kids can't find them," "a basket," "jar," "can," etc. One sympathizes with the mothers of the boys who apparently kept turtles and snakes in their drawers and boxes.

From question III, it is clear that the play purpose was the most frequent till twelve years old, with "keep it to look at," "keep it in case I need it," close seconds. After twelve, "keep it to look at" is much the most frequently checked by girls; that and "keep to show," "keep in case of need" about equally often checked by boys. Does this argue the "blindness" of the tendency which many adults betray when they rationalize the impulse to hoard what appears to be trash

by the excuse of liking to have it "for it might come in handy some day?"

The next table contrasts the purpose of play, and the purpose for use in school, for each sex separately, and for ages under twelve, twelve and over.

TABLE IX

	Under Twelve Years, Per Cent	Twelve Years, Per Cent	Twelve Years Plus, Per Cent
Play			
Boys	34	33	18
Girls	36	25	8
Use in School			
Boys	12	20	17
Girls	17	27	36

Here the serious purpose is seen at least increasing; though the high percentage in the case of the girls is doubtless influenced by the fallacy of selection of two hundred thirty girls of eighteen or more in high school or normal school. Reference to the list of items checked in Table VI will show the trivial nature of many of the collections, even in the older years. The purposes "keep to look at," "keep in case of need," might be either trivial or serious, depending rather on the nature of the collection. We have at least the contrast that, among the younger children, while eighty-nine per cent are collecting something or other, only about fifteen per cent are conscious of school guidance therein. Ninety-one per cent of the twelve year olds are collecting, less than a fourth of them connecting that with school. Ninety-five per cent of the older ones are collecting, about twenty-seven per cent for school use.

Conclusions

1. In the region where this questionnaire was used there is a general interest among children in collecting.

2. Whereas Burk considers the interest strongest in the ages from eight to eleven, these returns show (*a*) a greater median number of collections from nine to thirteen, (*b*) no special age of peak of interest judging by the number of collections, (*c*) an increasing strength judging by the number reporting affirmatively.

3. An itemized list such as ours, suggested by the form of the play quiz of Lehman and Witty, is probably more stimulating to

the memory—we hope not the imagination—of the subjects, than the indefinite form used by Burk.

4. Our returns corroborate Burk's in that interest in marbles, coupons, trivial things belongs chiefly to the younger years. Also that more boys than girls care for stamps, marbles, pieces of metal; more girls care for pieces of cloth, trimmings, ribbon, beads, and doll accessories.

5. Since during the school ages eight to fourteen so large a percentage of all children admitted activity in collecting and revealed a wide range of interests, it is worth while for educators to recognize and utilize this tendency.

AN EXPERIMENTAL STUDY OF BRIGHT AND DULL HIGH SCHOOL PUPILS

W. H. PYLE AND G. H. SNADDEN

Detroit Teachers College

The purpose of this study may be stated as an attempt to throw some additional light on the problem of the essential difference between bright and dull pupils. This is a continuation of a similar experiment made by Dr. E. L. Schott at the University of Missouri in connection with one of the authors[1] and published as *University of Missouri Bulletin*, Vol. XXXVI, No. 13. The study here reported differs from the former in that more extensive experiments have been made, especially in our comparison of the learning curves of the two groups.

Selection of Subjects

Our subjects were selected from the pupils of the New Central High School of Detroit. Our bright group consists of the thirty-one students having the highest IQ's and the best scholarship of all the students in the high school. All of this group were on the honor roll and ten of them had an all "A" record. Included in the group are the two girls and the two boys having the highest records for both intelligence and scholarship of all the pupils in the senior class. From this group of thirty-one pupils, we selected for the learning curve experiments the eight highest as is explained later.

The dull group, eight boys and four girls, had the lowest IQ's and the poorest scholarship records of all the pupils in the high school. All were failing in most of their studies, and the whole record of each of them had been one of failure.

The Experiments

There were two sets of experiments: (1) A group of nine ideational experiments and (2) a group of three motor learning experiments and a series of muscular speed tests. The extensive, individual motor learning experiments, from which the learning curves were constructed, were given only to the dull group and the eight of the bright group who had made the highest records in the group of ideational

[1] W. H. P.

experiments. The ideational learning experiments were as follows: Visual learning, auditory learning, reasoning, analogies, army alpha test, literary interpretation test, mental multiplication test, key-alphabet test, and associative coupling test. The motor learning tests were the substitution experiment, the card-sorting experiment, the manthanometer experiment, and a series of muscular speed tests with the tapping board, these tests were each of thirty seconds duration. There were twenty successive experiments in the motor learning tests, with identical procedures for every subject.

The Results

The results are shown in a comparative manner in the accompanying figures. There are two important points to notice in studying the figures. One is the amount of overlapping in the two groups as shown in Charts I and IX. The other point is the superiority of the best in the poor group as compared with the poorest in the bright group, as shown in Charts III, V, and VII.

In the ideational types of experiments, it will be seen that there is no overlapping of the groups in six of the nine experiments. But there is overlapping in three of the experiments. In the learning-

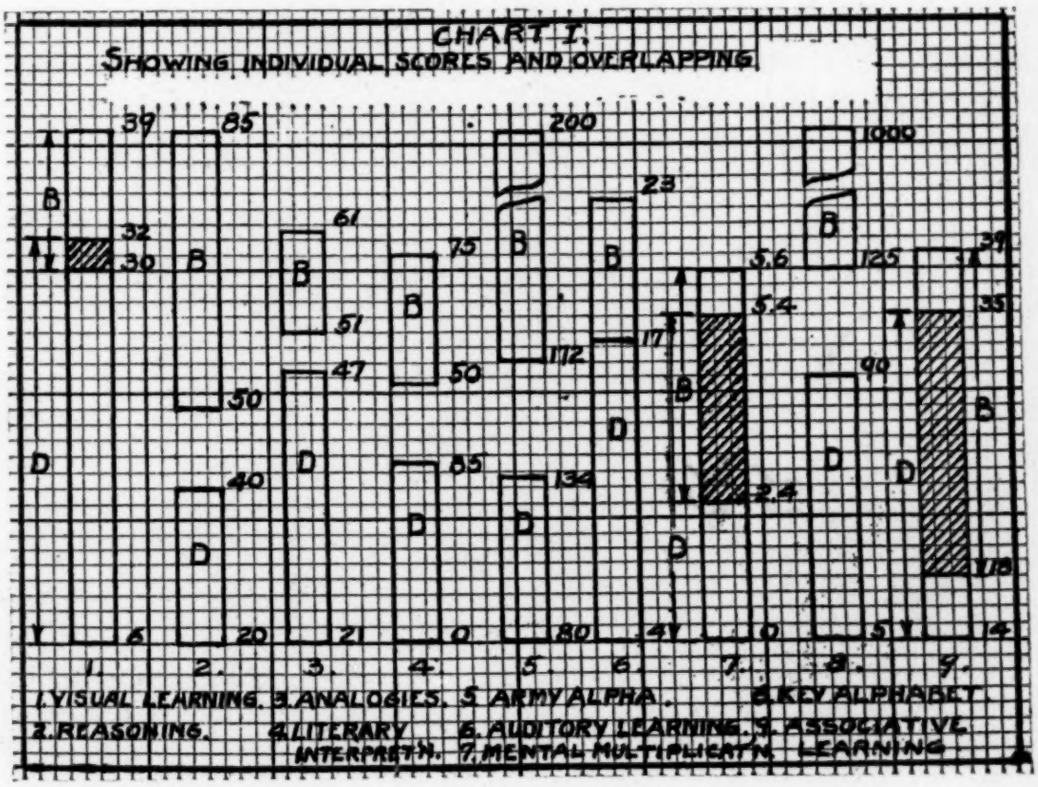
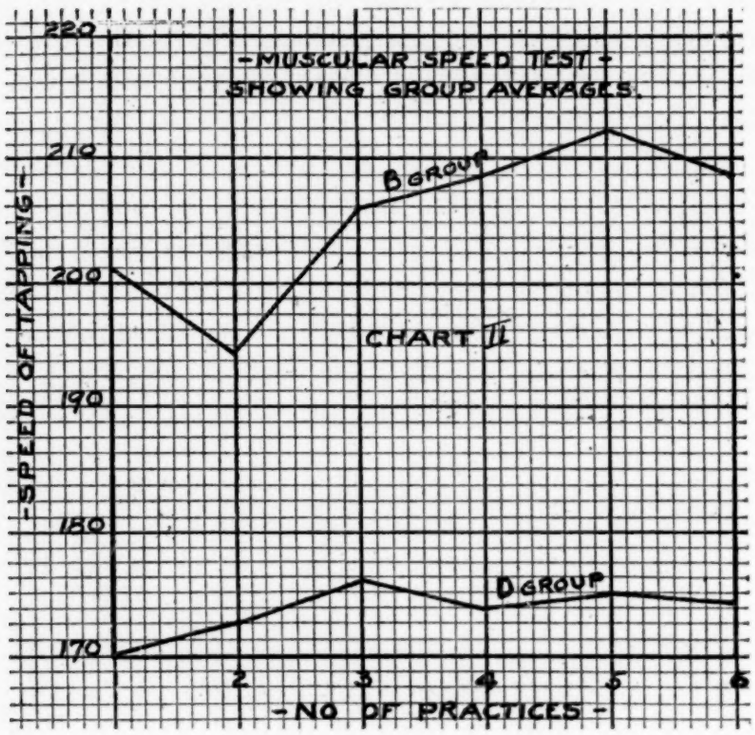
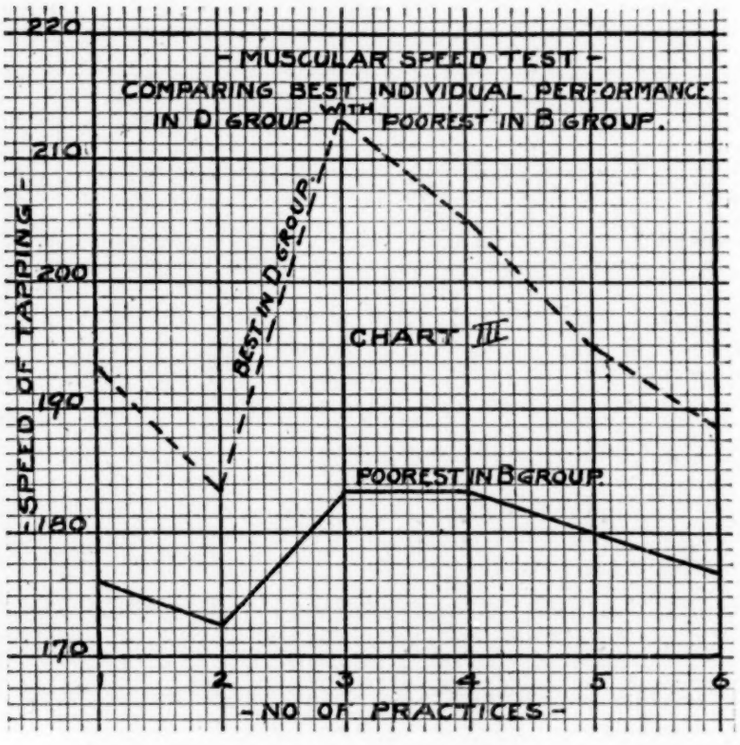
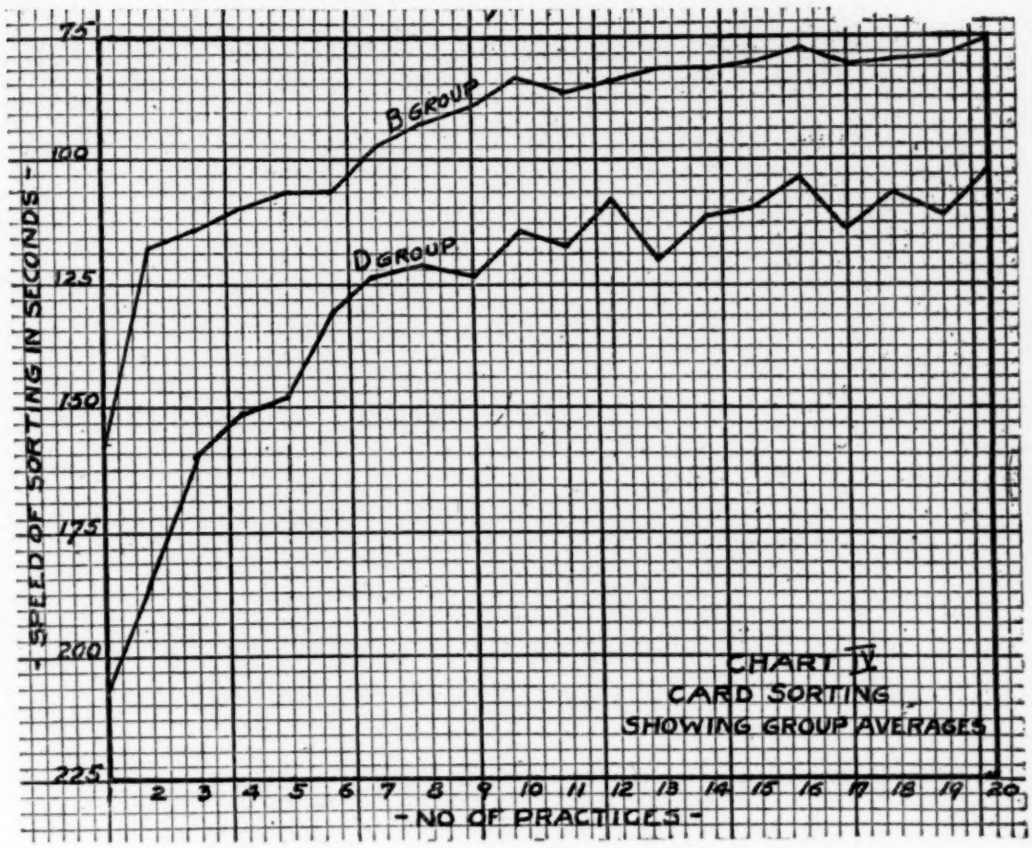
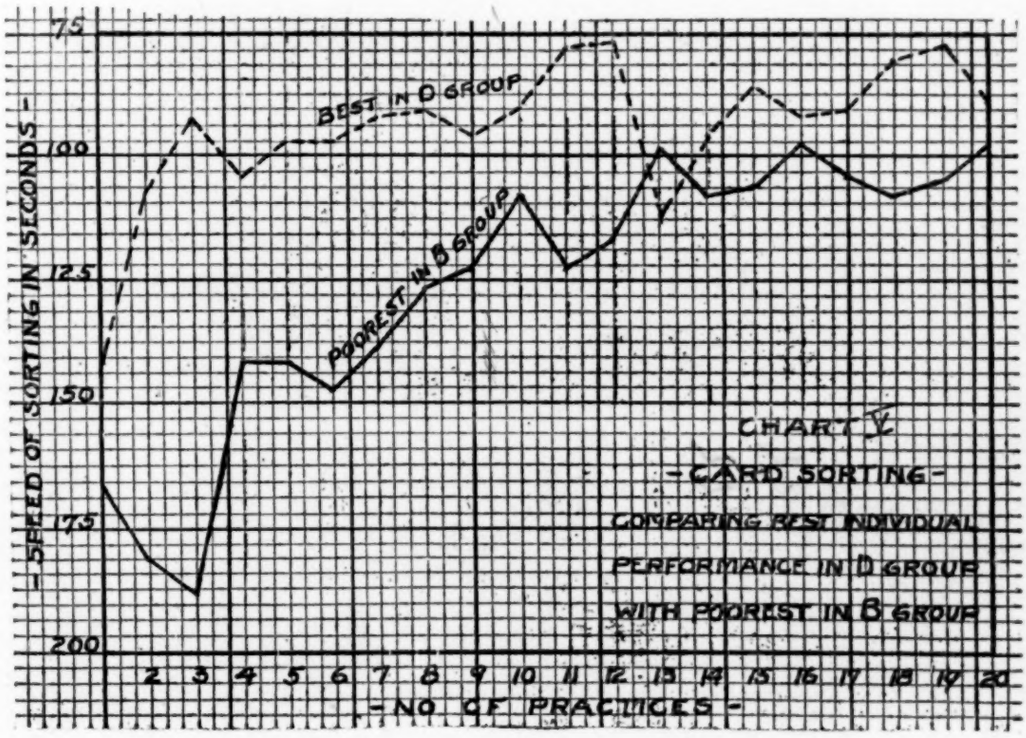


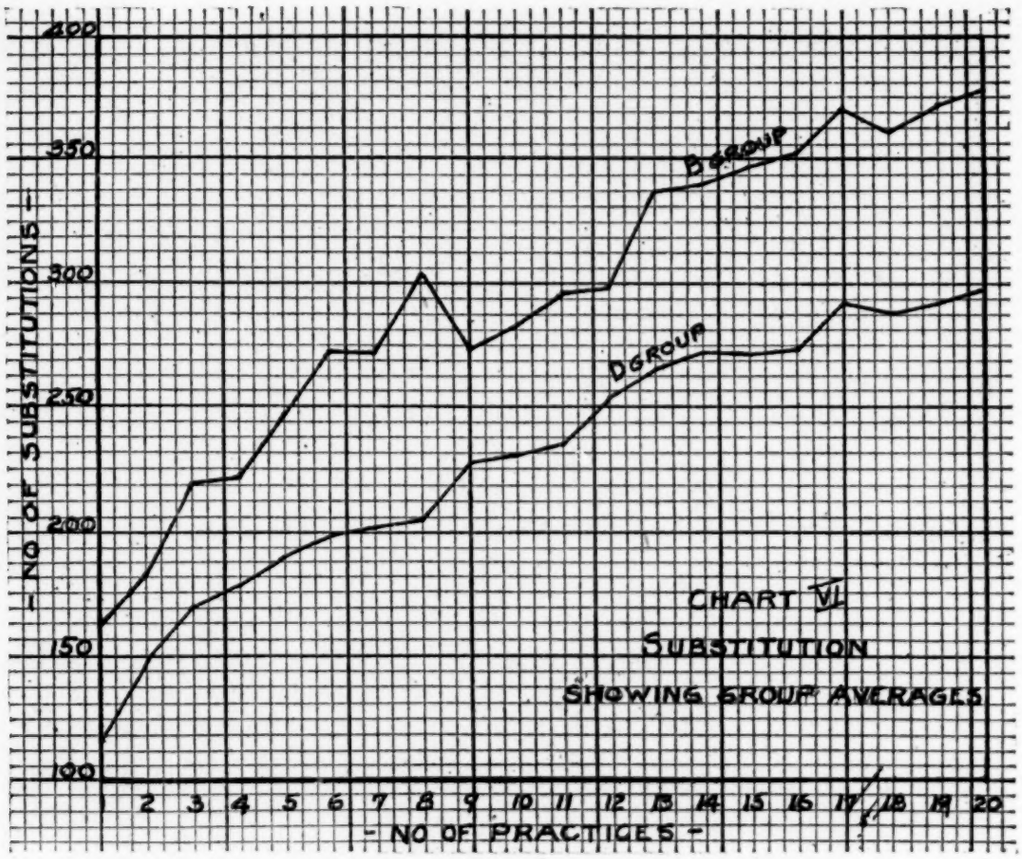
CHART I.
SHOWING INDIVIDUAL SCORES AND OVERLAPPING
1. VISUAL LEARNING. 3. ANALOGIES. 5. ARMY ALPHA. 8. KEY ALPHABET.
2. REASONING. 4. LITERARY INTERPRETN. 6. AUDITORY LEARNING. 7. MENTAL MULTIPLICATN. 9. ASSOCIATIVE LEARNING

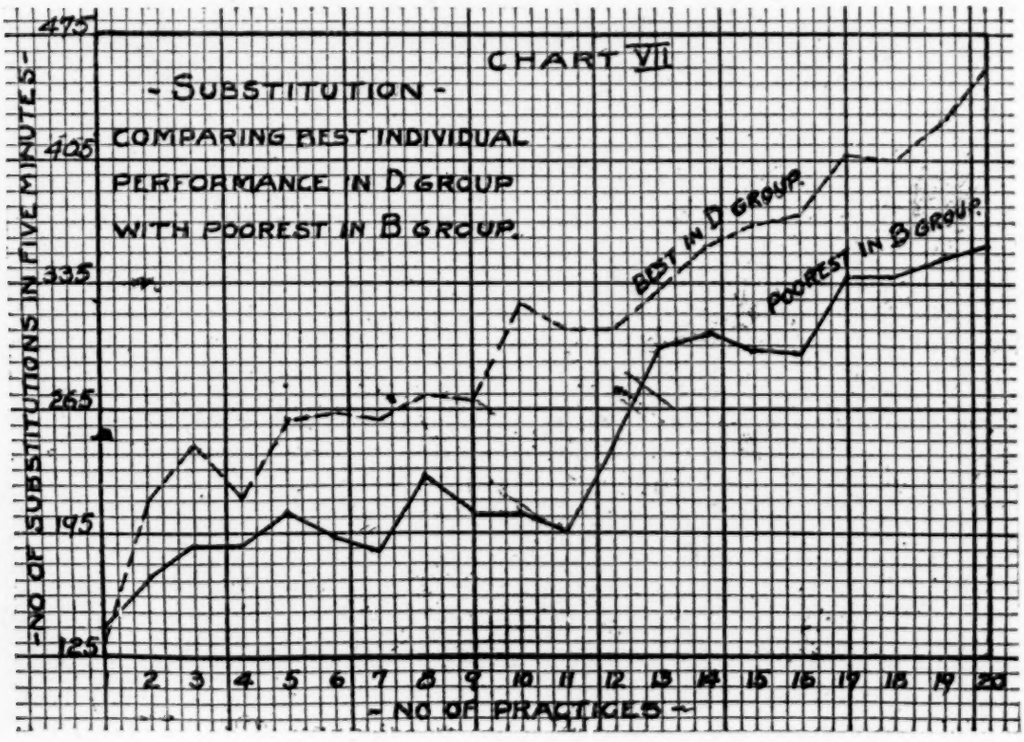

-MUSCULAR SPEED TEST-
SHOWING GROUP AVERAGES.
B GROUP
D GROUP
CHART II
-SPEED OF TAPPING-
220
210
200
190
180
170
1
2
3
4
5
6
-NO OF PRACTICES-


-MUSCULAR SPEED TEST-
COMPARING BEST INDIVIDUAL PERFORMANCE
IN D GROUP WITH POOREST IN B GROUP.
BEST IN D GROUP
POOREST IN B GROUP
CHART III
-SPEED OF TAPPING-
220
210
200
190
180
170
1
2
3
4
5
6
-NO OF PRACTICES-


75
100
125
150
175
200
225
B GROUP
D GROUP
- SPEED OF SORTING IN SECONDS -
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
- NO OF PRACTICES -
CHART IV.
CARD SORTING
SHOWING GROUP AVERAGES


75
100
125
150
175
200
BEST IN D GROUP
POOREST IN B GROUP
- SPEED OF SORTING IN SECONDS -
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
- NO OF PRACTICES -
CHART V.
- CARD SORTING -
COMPARING BEST INDIVIDUAL
PERFORMANCE IN D GROUP
WITH POOREST IN B GROUP


400
350
300
250
200
150
100
- NO OF SUBSTITUTIONS -
B GROUP
D GROUP
CHART VI
SUBSTITUTION
SHOWING GROUP AVERAGES
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
- NO OF PRACTICES -


475
405
335
265
195
125
- NO OF SUBSTITUTIONS IN FIVE MINUTES -
CHART VII
- SUBSTITUTION -
COMPARING BEST INDIVIDUAL
PERFORMANCE IN D GROUP
WITH POOREST IN B GROUP.
BEST IN D GROUP.
POOREST IN B GROUP.
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
- NO OF PRACTICES -

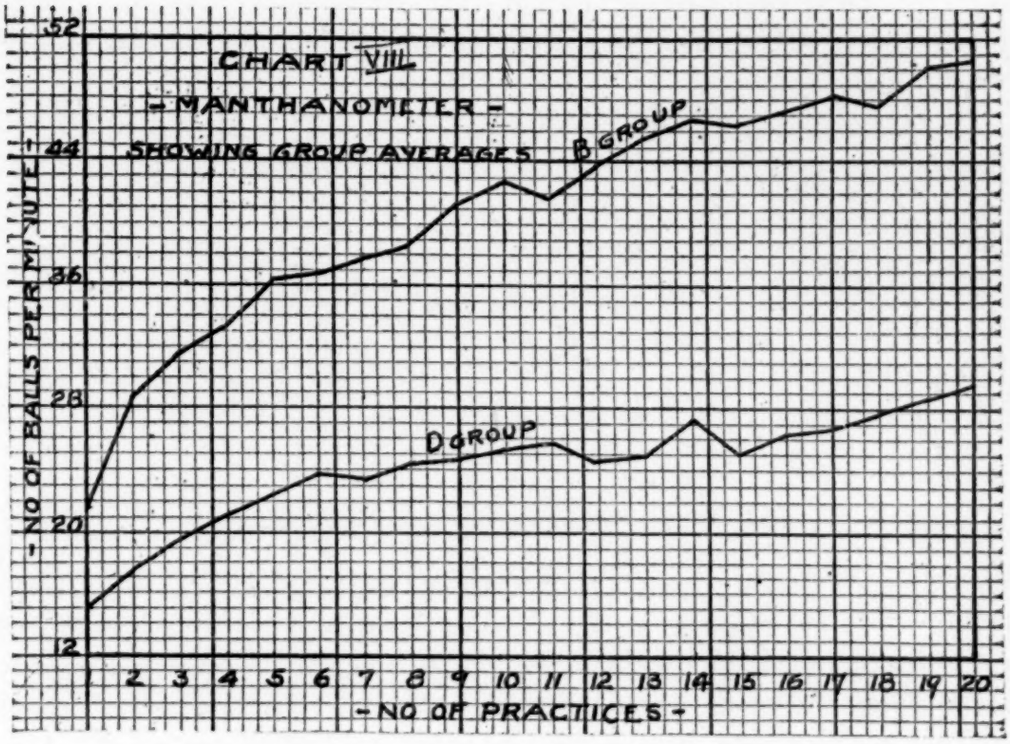


CHART VIII
- MANTHANOMETER -
SHOWING GROUP AVERAGES
B GROUP
D GROUP
- NO OF BALLS PER MINUTE -
52
44
36
28
20
12
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
- NO OF PRACTICES -

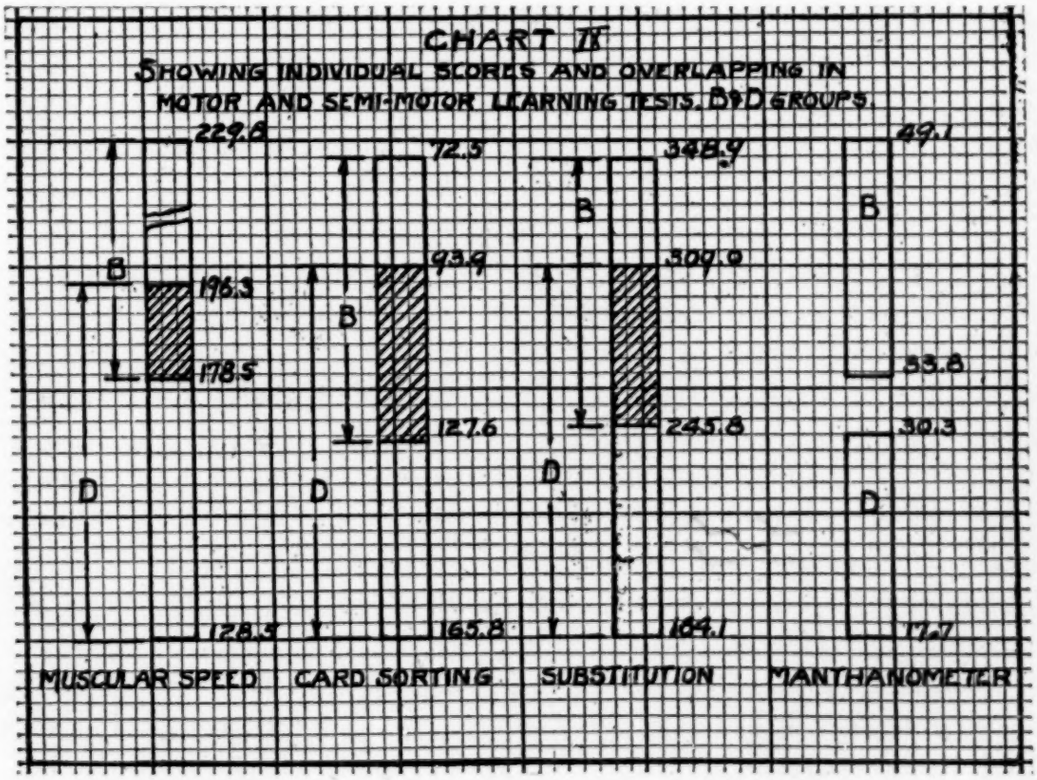


CHART IX
SHOWING INDIVIDUAL SCORES AND OVERLAPPING IN MOTOR AND SEMI-MOTOR LEARNING TESTS. B & D GROUPS.
229.6
196.3
178.5
128.3
72.5
93.9
127.6
165.8
348.9
309.9
245.8
184.1
49.1
33.8
30.3
17.7
B
D
MUSCULAR SPEED
CARD SORTING
SUBSTITUTION
MANTHANOMETER

curve experiments and speed tests, the groups overlapped in muscular speed, card sorting and substitution, but not in the manthanometer experiment.

In Charts II to VIII, it is seen that although the groups are separated as groups on the basis of the averages, some individuals of the dull group excel some of those in the bright group. Of the three forms of motor learning, the manthanometer experiment is decidedly the most complicated and difficult, and in this experiment the two groups are completely separated. It may be that the more difficult and complicated is a form of learning, the more is the demand made upon some general factor of learning.

Interpretation of the Results

The net results of our experiments may be stated in very few words. If the brightest and most successful high school pupils are compared with the dullest and least successful pupils, it is found that in some types of work including certain forms of learning, some members of the dull group will excel some members of the bright group. This means that the various abilities or functions of a person are not uniform. In the same person some functions will be possessed in high degree and some in low degree. One of the brightest pupils in a high school may be excelled in some forms of learning by a pupil rated low in intelligence (as determined by an intelligence test), and who is a failure as a high school student in the ordinary academic studies.

For practical purposes, therefore, we should be cautious in the value we put upon a single type of measure of a person's ability or even from a pooling of the results of different types of measures. Averaging the mental traits of an individual is perhaps as foolish as the averaging of his physical traits would be. What, for example, would it mean to average the measures of a person's height, weight, lung capacity, cephalic index, eye-color, hair-color, and blood pressure? Nothing, of course. These physical traits are found in all sorts of combinations. So, it seems, are different mental traits found in a great variety of combinations.

Whatever may be the facts concerning a central or general factor of learning capacity, specific abilities, as they express themselves when submitted to experimentation, may vary widely in the same individual.

Explanation of the Charts

The bright group is indicated by "B;" the dull group, by "D."

Chart I shows the range of scores made by the members of the bright and dull groups in the nine ideational experiments. Charts II, IV, VI, and VIII show the average standing of the bright and dull groups in the motor learning experiments and the speed tests. Charts III, V, and VII compare the curves of the best in the dull group with the poorest in the bright group.

ARE BOYS EXCELLING GIRLS IN GEOMETRIC LEARNING?

WINONA M. PERRY

Associate Professor of Educational Psychology and Measurements, University of Nebraska

Are boys excelling girls in geometric learning? This question was suggested by several recent studies in this general field, sex differences in achievement in geometry. In 1916 Stockard and Bell[1] found that boys slightly exceeded girls in their responses to a given test. In 1919 Touton[2] studied the examination papers of boys and girls to determine those abilities which are essential in the solving of geometric problems. The number of students, 2800, was at least sufficient; but the selection of the boys and girls who took this Regents' Board Examination was from those students who had reached the higher levels of achievement in their classes in geometry. In the most recent study, Webb[3] concluded that superior geometric ability was associated with boys as a group and especially with those boys of the four lower mental age levels. His measure of achievement was the scores made by boys and by girls upon the Webb Geometry Test. In this test the major emphases of Parts I–IV, were the meanings of certain terms and the conditions under which triangles could be proved equal. In Part V, achievement in construction problems was measured in terms of correct answers which were the results or products of thinking. The student responses to situations indicated that these tests did not detect the adequate from the inadequate responses within a partially correct solution of an exercise. A more diagnostic instrument would reveal the factors or the more unit connections within the process of solving the separate problem situations.

In 1924–1925 an investigation was conducted in the Charles E. Gorton High School in Yonkers, New York, by the writer as a member of the staff of the Lincoln School of Teachers College. There were three divisions of students: The experimental division (II) and a

[1] Stockard, L. V. and Bell, J. C.: A Preliminary Study of the Measuring of Abilities in Geometry. *Journal of Educational Psychology*, Vol. VII, 1916, pp. 567–580.

[2] Touton, Frank C.: "Solving Geometric Originals." Teachers College Contributions to Education No. 146, 1924.

[3] Webb, Paul E.: A Study of Geometric Abilities among Boys and Girls of Equal Mental Ability. *Journal of Educational Research*, Vol. XV, 1927, pp. 256–262.

control division (I) were taught by the writer; while the teacher of the additional control division (III) was a regular member of the high school staff. Through contrasting the responses of students in the experimental division with those in the control divisions we reached an answer to the question, Can students be taught to attack the solution of exercises in Plane Geometry with the assurance either that they can be successful during the first trial, or, if unsuccessful in the first attempt, can they retrace their steps only to that part necessary in order to redirect their thinking? This question was answered affirmatively in "The Study of the Psychology of Learning in Geometry."[1] However, the technique used toward solution of geometry originals was differential in its effect, as it proved to be of "marked benefit to the group of average students, of slight benefit to the group of good students, and detrimental to the group of superior students in initiating and developing habits of analysis and generalizations necessary for successful solution of these originals." I have examined the data from this investigation to determine whether the boys were equal, superior, or inferior to the girls in the solving of certain geometry originals.

The Personnel of These Divisions

The number of boys and of girls in the control divisions (I and III) and in the experimental division (II), also in the ability groups are presented in Table I.

Table I.—Sex Distribution in Divisions and in Ability Groups

Group	Division I (control)			Division II (experimental)			Division III (control)			Divisions I, II, III
	B	G	Total	B	G	Total	B	G	Total	
S.....	3	2	5	3	2	5	4	1	5	15
Gd.....	5	4	9	5	4	9	6	3	9	27
Av.....	8	7	15	8	8	16	11	5	16	47
Total B16...	..	..	..	16	..	..	21	..	..	53 or 59.55 per cent
Total G13...	..	13	..	..	14	..	..	9	..	36 or 40.45 per cent
Grand total...	..	..	29	..	..	30	..	..	30	89

Thus, approximately sixty per cent of the total number of students were boys and forty per cent were girls. The initial and final abilities

[1] Perry, W. M.: "A Study in the Psychology of Learning in Geometry." Teachers College Contributions to Education No. 179, 1925.

of the boys and of the girls to reason were determined by a "Composite Test in Reasoning," Forms A and B. Their general mental ability was measured by the Terman Group Test of Mental Ability, Form B. The median scores of the boys and girls upon these tests with the frequency and per cent of boys reaching or exceeding the median girl scores are found in Tables II and III.

TABLE II.—MEDIANS—COMPOSITE TEST IN REASONING. FORMS A AND B—WITH PER CENT OF BOYS REACHING OR EXCEEDING THE MEDIAN GIRL

Division	Form A		Form B	
	B	G	B	G
I	255.00	250.00	330.00	307.50
II	250.00	245.00	340.00	335.00
III	258.75	258.75	353.75	353.33
Total	255.83	250.00	346.87	335.00
Overlapping	**F**	**Per cent**	**F**	**Per cent**
$B_I - G_I$	9	56.2	13	81.2
$B_{II} - G_{II}$	9	56.2	9	56.2
$B_{III} - G_{III}$	11.1	52.8	10.67	50.8
$B_T - G_T$	29	54.7	31	58.4

TABLE III.—MEDIANS—CA'S, MA'S, IQ'S—WITH PER CENT OF BOYS REACHING OR EXCEEDING THE MEDIAN GIRL

Median	CA		MA		IQ	
	B	G	B	G	B	G
I	14–8.5	14–6.5	16–10.5	15–6.25	110.0	107.5
II	15–0.5	15–1.5	16–7.0	15–4.7	112.0	107.0
III	15–1.5	15–6.5	16–6.5	16–7.25	110.5	106.5
Total	14–10.5	15–0.5	16–7.25	15–7.00	111.2	107.0
Overlapping						
$B_I - G_I$	10	62.5	12	75.0	9	56.2
$B_{II} - G_{II}$	8	50.0	14	87.5	10	62.5
$B_{III} - G_{III}$	5.5	26.1	10	47.6	13.5	64.2
$B_T - G_T$	24	45.2	40	76.6	32	60.3

As the median scores of the boys are slightly higher than the median scores of the girls and as the per cents of overlapping are also

slightly greater than fifty, the boys excel the girls somewhat in initial and in final reasoning ability. In chronological age the boys in division III were somewhat younger, in division II approximately of the same age, and in division I slightly older than the girls. In mental age the boys, with the exception of those in division III, exceed the girls. Furthermore the boys exceed the girls in IQ's also. Thus as a group the boys were slightly superior to the girls in ability to learn as well as in reasoning ability. Classifications of these students ranged from first semester, Grade IX to second semester, Grade XI; the median (simple) in each division was first semester, Grade X. It was found that the boys studied much less but were somewhat more regular in attendance than the girls.

Achievement in Tests in Exercises

There were three tests in exercises, each including three originals. The median scores of boys and of girls and the per cent of boys reaching or exceeding the median girl scores are given in Table IV.

Table IV.—Medians—Tests in Exercises with Per Cent of Boys Reaching or Exceeding the Median Girl

Median	Test 1		Test 2		Test 3	
	B	G	B	G	B	G
I	59.50	60.50	38.00	51.50	61.50	55.50
II	66.50	77.00	65.00	83.00	72.00	70.00
III	53.50	51.50	42.50	39.75	55.25	57.50
Total	55.83	60.00	44.50	63.00	59.50	58.50
Overlapping	**F**	**Per cent**	**F**	**Per cent**	**F**	**Per cent**
$B_I - G_I$	7.5	46.8	7	43.7	10	62.5
$B_{II} - G_{II}$	3	18.7	4	25.0	9	56.2
$B_{III} - G_{III}$	11.5	54.7	12.5	59.5	8	38.0
$B_T - G_T$	24	45.2	20	37.7	27	50.9

In three instances (two in division III and one in division II) the boys excelled the girls. With these exceptions, on each test in each division the girls excelled the boys in their ability to solve exercises in geometry. The extent of overlapping of the boys' scores beyond the median girl score is indicated by the per cents ranging from 18.7 to 62.5. Thus, although the boys were superior initially in

general mental ability and in reasoning ability, only in exceptional instances did their achievement exceed that of the girls. In general, then, the achievement of the girls was superior to that of the boys.

Association between Sex and Adequate Responses

In a further attempt to find out which group, the boys or the girls, responded the more adequately to these exercises in geometry, also to determine whether the correct responses were dependent upon the nature of the steps, upon the ability of the students in reasoning, or upon the division of which the students were members, I have used the technique described by Yule in his text "An Introduction to the Theory of Statistics."[1] Thus we may compare the number of correct responses made by the boys, $\frac{(AB)}{(A)}$, with the number of correct responses of the girls, $\frac{(\alpha B)}{(\alpha)}$. The question as to which responds the more adequately, the boys or the girls, may be determined by the following formula:[2]

$$\frac{(AB)}{(A)} \gtreqless \frac{(\alpha B)}{(\alpha)}.$$

This is a most searching means of finding out the degree of association between sex and adequate responses. If the per cent $\frac{(AB)}{A}$ is greater than, is equal to, or is less than the per cent $\frac{(\alpha B)}{(\alpha)}$, there results a "+," a "0," or a "−" difference which is indicative that the boys are excelling the girls in ability, that the ability of both sexes is equivalent, or that the girls are excelling the boys in ability respectively.

The steps "G" or "L" refer to the given or link steps respectively in each of the exercises attempted by these students. A "given" step means a step in the proof which is one of those steps directly gleaned from the statement of the problem, although expressed in terms of the figure. A "link" step, however, results from a more complicated process of thinking. It "is deduced through inference from the statement, or resulting suggestion from the figure."[3]

---

[1] Yule, G. U.: "An Introduction to the Theory of Statistics." London Griffin, 1922, 25ff.

[2] (A), (α), (B), (β) represent the number of boys, of girls, of correct responses, and of incorrect or omitted responses respectively.

[3] Perry, W. M.: *Op. cit.*, p. 13.

An illustration may make this point clearer. If the base of an isosceles triangle is divided into three equal parts the segments joining the points of division to the vertex are equal. The steps stating that two pairs of sides are equal are "given" steps. The step stating that the angles included between these sides are equal is a "link" step, as it is that part which allows the student to bridge the gap from given data to a conclusion desired. The divisions are indicated by I, II, and III; the ability groups, by S (Superior), Gd (Good), and Av (Average).

The actual numbers of correct boy responses (AB) and the correct girl responses (αB) to all of the steps irrespective of their nature are five hundred seventy-three and four hundred forty respectively. The possible numbers of correct boy responses (A) and of girl responses (α) one thousand two hundred seventy-two and eight hundred sixty-four respectively.

$$\therefore \frac{(AB)}{(A)} < \frac{(\alpha B)}{(\alpha)}$$

$$\therefore \text{D (or "difference") is } -.059$$

As this difference is a minus one, in general the girls excelled the boys slightly in responding to steps within the exercises.

This conclusion becomes the stimulus toward more detailed study of the responses in order to determine whether this somewhat greater achievement of the girls will be found equally among the responses to the given and to the link steps, by the students in each of the divisions, and by each of the ability groups. In Table V there are stated the number of boys and girls in the several divisions and in the ability groups with the actual differences in per cent of boys and of girls stating given or link steps correctly.

TABLE V.—NUMBER AND DIFFERENCES IN PER CENT OF BOYS AND OF GIRLS STATING GIVEN (0–7) AND LINK (0–17) STEPS CORRECTLY

	Given		Link		Differences	
	B	G	B	G	G	L
Total	259	175	314	265	.00+	−.08
I	73	61	100	82	−.01	−.34
II	82	71	125	138	.00+	−.12
III	104	43	89	45	.02	−.04
S	55	25	87	55	.07	−.01
Gd	83	57	86	85	.00+	−.13
Av	121	93	141	125	−.02	−.06

As the difference between the proportions of boys and of girls in responding to the given steps is .00+, this ability of boys and girls is practically independent though tending slightly in favor of the boys. Similarly the link step difference, −.08, shows that the girls in general excelled the boys somewhat in this ability. The divisions and the ability groups have one positive, one zero, and one negative difference among the given steps indicating almost equivalent ability on the part of boys and girls in responding adequately to the given steps. But in the separate divisions and in ability groups the link step differences are in each case minus, and vary from −.01 to −.34. Consequently the girls, irrespective of their division or ability group, were superior to the boys in their ability to deduce inferences from given data in these exercises.

Further evidence may be concluded from the number and proportions of boys and of girls, in divisions and in ability groups, who stated a specific given or a specific link step correctly in each exercise in each test. The differences—positive, negative, zero, and omitted (which means that no boy or girl responded to the step in question)—according to the various combinations of divisions, ability groups, nature of the step (given or link) and total number of correct responses are found in Table VI.

TABLE VI.—FREQUENCY OF DIFFERENCES (+, −, 0, OMITTED) IN WHICH BOYS OR GIRLS EXCEED IN SOLUTION OF GIVEN OR OF LINK STEPS

D	I			II			III			Ability			Total	Division		
	S	Gd	Av.	S	Gd	Av.	S	Gd	Av.	S	Gd	Av.		I	II	III
+	9	6	9	6	9	2	11	5	16	26	20	27	73	24	17	32
−	4	14	13	11	13	15	5	15	8	20	42	36	98	31	39	28
0	11	1	0	6	2	6	1	2	0	18	5	6	29	12	14	3
	0	3	2	1	0	1	7	2	0	8	5	3	16	5	2	9

D	S		Gd		Av.		Step		I		II		III	
	G	L	G	L	G	L	G	L	G	L	G	L	G	L
+	8	18	8	12	9	18	25	48	8	16	6	11	11	21
−	5	15	9	33	8	28	22	776	9	22	5	34	8	20
0	8	10	4	1	4	2	16	13	4	8	10	4	2	1
	0	8	0	5	0	3	0	16	0	5	0	2	0	9

In order to perceive more clearly the degree of association between sex and adequacy of responses one-half of the number of zero differences was added to both the number of positive and the number of negative differences. The ratios of the resulting numbers of plus and minus differences are found in Table VII.

TABLE VII.—BOY-GIRL RATIOS (+/−)
(Based on Table VI)

Group	I		II		III		G or L		Division	G or L		Total	
	+/−	10/x	+/−	10/x	+/−	10/x	+/−	10/x		+/−	10/x	+/−	10/x
S.....	1.52	6	.64	15	2.09	4	1.20	8	I	.81	12	.77	13
Gd....	.44	22	.71	14	.37	27	.50	20	II	.52	19		
Av....	.69	14	.27	37	2.00	5	.76	13	III	1.13	8		

From these boy-girl ratios, for every ten steps stated correctly by boys there was computed the approximate number of steps which girls could state correctly. Among these proportions eleven of the sixteen favor the girls. Thus the girls as a group, also in division II and in the "Good" and "Average" (except in III) ability groups, excel the boys when responding to the separate steps within these originals.

The amounts of the "+" and "−" differences are indicated by the median differences appearing in Table VIII. Among the given steps five favor the boys but among the link steps six favor the girls.

TABLE VIII.—MEDIAN DIFFERENCES (+, −)
(Based on Table VI)

D	S		Gd		Av.		Total		I		II		III	
	G	L	G	L	G	L	G	L	G	L	G	L	G	L
+	.50	.26	.25	.23	.24	.14	.34	.23	.50	.31	.20	.18	.34	.19
−	.33	.34	.21	.25	.25	.20	.26	.26	.34	.19	.20	.26	.22	.34

Similarly, from the total number of boy and girl differences in responding to either the given or the link steps we find the per cents of difference favor the boys somewhat if the step is "given," but rather markedly favor the girls if the step is "link." Thus the superiority in achievement is associated with boys to a slight degree only if the

steps are given ones but with girls more markedly if the steps are "link" in nature. Although considering the steps irrespective of their nature, the superiority in responding adequately to them in the different originals is somewhat more strongly associated with the girls.

A further question arises: Will the girls maintain their superior achievement in responding adequately to combinations of steps within an exercise, rather than to any separate step within a single exercise? There was found[1] the number of boys and girls who stated correctly the steps, given or link, in an entire exercise in each of the tests, also the number of boys and girls who stated correctly the given and link steps in all of the exercises. More boys than girls were unable to respond adequately to a single exercise as a whole. Also more boys than girls responded correctly to one, two, three, and to eight exercise combinations of steps. Two boys and two girls solved seven exercise combinations correctly. More girls than boys solved four, five, or six exercise combinations correctly. Thus girls were somewhat superior to boys in responding adequately to the total step combinations within the separate exercises.

Furthermore, the number of plus, minus, zero or omitted differences proved to be a further measure of the superiority of boys or of girls who stated the separate exercise combination of steps correctly. Also, the boy-girl ratios indicate that the girls as a group, and the girls in the two upper ability groups and in divisions I and II were superior in this trait. The median differences are "−" differences, except in the "Average" group (+.04) and in division III (+.17). The amounts of differences (as shown by the median "+" or "−" differences) favor each sex equally or one sex more than the other equally often.

The differences in the proportions of boys and girls stating these numbers of step combinations correctly within a single exercise and within each test are indicating either equality in achievement (as one plus, one minus, and one zero difference occurs when three exercises were stated correctly) or superiority on the part of the girls. As there were three plus differences for zero exercises stated correctly, these are indicative of more boys than girls being unable to respond correctly to an entire exercise. Thus when considering the step combinations within exercises and tests, girls proved to be superior in the solutions of these geometric exercises.

[1] Certain basic tables have not been printed, due to lack of space. Upon request, however, the interested reader may secure these from the author.—W. M. P.

Résumé of This Study

In the earlier investigations of Touton and of Webb, the conditions of their experiments differed from those of this study in certain respects. The boys and girls whose test papers Touton examined in detail were among the more able students from the classes in plane geometry in the state of New York. The Webb Test revealed certain abilities developed in classes in Plane Geometry, but it does not seem to be sufficiently diagnostic of those nearer unit abilities constituent within the process of the solution of exercises in geometry. Although the numbers of boys and girls were comparatively large in each of these earlier studies, this investigation included all of the students in three classes at least somewhat characteristic of high school classes in Plane Geometry conducted by teachers in the United States. As the personnel of the ability groups, of the divisions and of the sexes numbered from one to fifty-three, in addition to the method of measuring the per cent of boys' scores reaching or exceeding the median girl score, a most searching technique was used in order to ferret out those rather subtle factors operative during the actual process of solving exercises in geometry. As a result, we were led to conclusions differing from those drawn from these earlier studies.

Although as a group the boys were somewhat higher in general mental ability and in reasoning ability, the girls reached a higher degree of achievement in the solution of exercises in geometry. The degree of superiority in achievement depended less upon the ability group and the division membership of these students but more markedly upon the nature of the step. Adequate responses to the "given" steps were found nearly as frequently among girls as boys, though tending slightly in favor of the boys. Ability to respond adequately to the "link" steps, however, was more characteristic of the girls. This ability led to the somewhat higher achievement, also on the part of the girls, in responding the more adequately to the step combinations within the several exercises. Therefore it proved to be the girls who were the more superior in responding more adequately to the solutions of exercises in geometry. Especially were the girls the more superior in responding to those "link" steps, by means of which they were able to bridge the gap between the given relations (those observed with the pertinent recalled facts) and the required relations which had controlled the direction of the processes of thinking during the actual solving of the exercises in geometry.

ERRONEOUS FIRST IMPRESSIONS ON OBJECTIVE TESTS

C. O. MATHEWS

Ohio Wesleyan University

"One reason for your poor showing on the true-false test is that you made too many changes after once recording a response," an instructor was recently heard to tell one of his students. He went on to suggest that in taking such tests she should have more faith in her first impressions.

About the same time that these statements were overheard, the writer noticed the following statements in a book recently published in regard to taking true-false tests:

> Some students and teachers stoutly maintain that, in a true-false test, if you stop to think you are more likely than not to make the wrong answer; that your first reaction is more reliable than second thought. Some unpublished data collected under the writer's direction reveal, however, that when persons have second thoughts on true-false tests and change their answers accordingly, they raise their grades about twice as often as they lower them.[1]

It was to verify Crawford's statement and to furnish data which would indicate how students should be instructed in taking objective tests that this study was planned.

Students' Opinions Regarding Changes

Do students think that such changes usually result in correct answers, or are they of the opinion that it is unprofitable to check over carefully such tests before handing them in?

Soon after the close of a semester, opinions were sought from fifty-one students selected from those who had just completed the course in educational measurements or educational psychology. They were selected so as to be representative of the various degrees of achievement in these classes. Each student had taken at least four one-hour new-type tests during the course. The following letter was sent to each of these students:

---

[1] Crawford, Claude C.: "The Technique of Study." Houghton Mifflin Co., 1928, p. 277.

My dear ————,

I am working on some materials related to new-type objective tests. Since you took a number of these tests in one of my courses, I am writing to ask if you will favor me by replying on the enclosed postal card to the question in the next paragraph:

In taking true-false and multiple-choice types of tests one often marks an answer and then, after further consideration, changes his response. Sometimes these changes result in the wrong answer being indicated last, and at other times the right answer is indicated last. People differ in this respect. *In your own personal case, do you think such changes more often resulted in final wrong answers or in final right answers?*

I want to thank you in advance for this favor. I will value your reply.

Yours sincerely,

The postal card contained only the following statement and a place for the student's name: "I think such changes more often resulted in final ———— answers for me."

Of the fifty-one students twenty-eight replied. Twenty-four expressed the belief that their changes more often resulted in wrong answers; three thought that the right answer was the more frequent and one said he thought that in the long run, half of the changes tended to be each way, thus not affecting the score.

A large majority of these students evidently believed that they did not profit from the changes which they had made in these examinations. If their opinions are correct then first impressions on these tests yielded higher scores. If their opinions are contrary to the facts, then they have a right to know it and should be informed prior to taking such tests of the value of making careful changes.

Analysis of Responses on Tests

In order to determine the facts relative to these points an analysis was made of the changed responses on a number of objective tests given to certain college classes during the year 1927–1928. There were tests from two classes in educational measurements, two in principles of teaching, three in educational psychology and one in educational research. The tests had been given as measures of achievement in these various subjects in the regular course of the work. The tests differed in length but time limits under which they were given were such that practically all students had finished prior to the end of the period. This allowed adequate time to go back over the questionable items, making changes when further deliberation

ERRONEOUS FIRST IMPRESSIONS ON OBJECTIVE TESTS

C. O. MATHEWS

Ohio Wesleyan University

"One reason for your poor showing on the true-false test is that you made too many changes after once recording a response," an instructor was recently heard to tell one of his students. He went on to suggest that in taking such tests she should have more faith in her first impressions.

About the same time that these statements were overheard, the writer noticed the following statements in a book recently published in regard to taking true-false tests:

> Some students and teachers stoutly maintain that, in a true-false test, if you stop to think you are more likely than not to make the wrong answer; that your first reaction is more reliable than second thought. Some unpublished data collected under the writer's direction reveal, however, that when persons have second thoughts on true-false tests and change their answers accordingly, they raise their grades about twice as often as they lower them.[1]

It was to verify Crawford's statement and to furnish data which would indicate how students should be instructed in taking objective tests that this study was planned.

Students' Opinions Regarding Changes

Do students think that such changes usually result in correct answers, or are they of the opinion that it is unprofitable to check over carefully such tests before handing them in?

Soon after the close of a semester, opinions were sought from fifty-one students selected from those who had just completed the course in educational measurements or educational psychology. They were selected so as to be representative of the various degrees of achievement in these classes. Each student had taken at least four one-hour new-type tests during the course. The following letter was sent to each of these students:

---

[1] Crawford, Claude C.: "The Technique of Study." Houghton Mifflin Co., 1928, p. 277.

My dear ————,

I am working on some materials related to new-type objective tests. Since you took a number of these tests in one of my courses, I am writing to ask if you will favor me by replying on the enclosed postal card to the question in the next paragraph:

In taking true-false and multiple-choice types of tests one often marks an answer and then, after further consideration, changes his response. Sometimes these changes result in the wrong answer being indicated last, and at other times the right answer is indicated last. People differ in this respect. *In your own personal case, do you think such changes more often resulted in final wrong answers or in final right answers?*

I want to thank you in advance for this favor. I will value your reply.

Yours sincerely,

The postal card contained only the following statement and a place for the student's name: "I think such changes more often resulted in final ———— answers for me."

Of the fifty-one students twenty-eight replied. Twenty-four expressed the belief that their changes more often resulted in wrong answers; three thought that the right answer was the more frequent and one said he thought that in the long run, half of the changes tended to be each way, thus not affecting the score.

A large majority of these students evidently believed that they did not profit from the changes which they had made in these examinations. If their opinions are correct then first impressions on these tests yielded higher scores. If their opinions are contrary to the facts, then they have a right to know it and should be informed prior to taking such tests of the value of making careful changes.

Analysis of Responses on Tests

In order to determine the facts relative to these points an analysis was made of the changed responses on a number of objective tests given to certain college classes during the year 1927–1928. There were tests from two classes in educational measurements, two in principles of teaching, three in educational psychology and one in educational research. The tests had been given as measures of achievement in these various subjects in the regular course of the work. The tests differed in length but time limits under which they were given were such that practically all students had finished prior to the end of the period. This allowed adequate time to go back over the questionable items, making changes when further deliberation

seemed to justify them. Students had been encouraged to make such changes with the suggestion that these changes were more apt to raise than lower their final scores.

The method of procedure was to go over these tests tabulating the items whose responses had been changed. It was impossible to detect many of the changes which students had spontaneously made because the items had not been arranged nor instructions given with this study specifically in view. For example, the directions on all true-false tests specified that statements which were true were to be indicated by a plus sign and those which were false by a minus sign. This probably prevented detection of at least half of the changes made in the true-false items, for items first marked minus could be changed to plus without any erasures or scratching out. Approximately half of the items on each test form were false.

Moreover, some of the changes made on items first marked plus could not be detected with certainty, because the first response had been totally erased or so thoroughly scratched that the order of change was doubtful. Only those items in which the order of changed responses was evident were included in the study.

Multiple-choice types, as the term is used here, include also a few matching and best answer tests. In practically all cases there were four or more alternatives in multiple-choice items. Each alternative was numbered and the directions requested the students to place the number of the best answer in parentheses. A number of the changes which pupils actually made could not be detected for the same reasons stated above in regard to the true-false items first marked plus. Only those items were included in which the order of change was beyond question.

The changes were tabulated under several categories which indicate the order of the responses made by pupils before turning in their papers to be scored. These categories and the number of items found under each are shown in Table I. The total number of items (*i.e.*, the total of all items of this type on tests examined) canvassed in order to locate these changes was 21,704 of the multiple-choice type and 17,990 of the true-false type.

In Table I *W* means a wrong response, *R* a right response, and *O* an item to which the response has been omitted. It will be seen that the changes are grouped according to types. Group I indicates items to which the first responses were wrong and the last responses were right. Group II is just the opposite of Group I. Group III

is made up of items to which the first responses were wrong and, although changes were made, the last responses were also wrong. In Group IV the first and last responses were right. Groups V and VI are items in which the responses were finally omitted, although the first response in one group were right and those of the other group were wrong.

TABLE I.—ORDER AND NUMBER OF CHANGES MADE BY COLLEGE STUDENTS IN RESPONSES TO TWO TYPES OF ITEMS

Group	Order of changes	Number of changes on each test type	
		Multiple-choice	True-false
I	*W-R*	287	340
	W-W-R	4	
II	*R-W*	115	158
	R-W-W	1	
	R-W-R-W	...	1
	R-W-R-W-R-W	...	1
III	*W-W*	135	
	W-W-W	2	
	W-R-W	...	3
	W-W-R-W	1	
IV	*R-W-R*	5	11
V	*R-O*	1	24
VI	*W-O*	3	31
	W-W-O	1	1
Total.....	.....	555	570

All the true-false tests were scored by subtracting the number of wrong responses from the number of right ones. Multiple-choice tests were scored on the basis of the number right. Students had been informed of the method of scoring prior to taking the tests. This probably accounts for the small number of omissions in the multiple-choice items after a response was once recorded. It also accounts for the fact that the changes in the true-false tests under under Group VI tended to raise the scores on the tests.

Tables II shows the extent to which the changes of responses raised, lowered, or left unchanged the final score on these tests. Here is found the answer to the principal question under investigation. It pays the student to go back over such objective test items as the true-false and multiple-choice types, making the changes which one's

best judgment can justify. If these figures are fairly representative, individuals who thus change their responses profit, on the average, much oftener than they lose. Fifty-three per cent of the changes on the multiple-choice items tended to raise the score as opposed to twenty-one per cent which lowered it. This is a difference of thirty-two per cent.

TABLE II.—THE NUMBER AND PER CENT OF CHANGES WHICH RAISED, LOWERED OR LEFT UNCHANGED THE SCORES ON TWO TYPES OF TESTS

Effect of score	Type of test					
	Multiple-choice			True-false		
	Groups	Number of items	Per cent	Groups	Number of items	Per cent
Raised	I	291	53	I, VI	372	63
Lowered	II, V	117	21	II, V	184	34
Left unchanged	III, IV, VI	147	26	III, IV	14	3
Total	.....	555	100	.....	570	100

Of the changes on true-false items, sixty-three per cent resulted in raising the score while thirty-four per cent lowered it. Here the difference in favor of raising the score is twenty-nine per cent. Both of these differences are statistically significant. The standard error of the difference in each instance is over ten times the difference.

INDIVIDUAL DIFFERENCES IN REGARD TO CHANGES

It is possible that there are individuals who *usually* lose when they change their first responses to test items. This study does not prove nor disprove the fact. The data used here did not yield enough changes for each student to make possible valid conclusions in regard to the tendencies of individuals.

There seems to be, however, fairly significant differences between superior and inferior students as to the number and kinds of spontaneous changes which they make on multiple-choice and true-false types of test items. Data were collected on this point by tabulating all discernible changes made by the thirteen best students and the thirteen poorest students in the classes in educational measurements and educational psychology.

The changes were tabulated for these two groups in the same manner as shown in Table I. These were then combined so as to show how the changes tended to affect the test scores. Table III shows a comparison of the two groups.

TABLE III.—DIFFERENCES BETWEEN THE PER CENTS OF CHANGES MADE ON TWO TYPES OF TESTS BY SUPERIOR AND INFERIOR COLLEGE STUDENTS

Effect on score	Superior students		Inferior students		Difference in per cent	Standard error of difference
	Number of changes	Per cent	Number of changes	Per cent		
Raised	53	67.9	43	49.4	18.5	2.43
Lowered	18	23.1	23	26.5	3.4	6.5
Left unchanged	7	9.0	21	24.1	15.1	5.6
Total	78	100.0	87	100.0		

These facts show that inferior students tended to make more changes than the superior ones. Yet, in spite of the larger number of changes, they made an appreciably smaller per cent of these changes in such a way as to raise their test scores. Of the changes which had no effect on the test scores, three times as many were made by inferior as by superior students.

Even though the inferior student seems to profit less than the superior one from the deliberation he gives to test items once marked, yet he tends to raise his score nearly twice as often as he lowers it by such changes. Therefore, he should be encouraged to go back over his tests and give further consideration to the more difficult and questionable items.

SUMMARY

1. Opinions among students and instructors vary as to the value of making deliberative changes on true-false and multiple-choice types of tests after a response once has been recorded. Twenty-four out of twenty-eight representative college students erroneously expressed the belief that when they made such changes they were more often wrong than right.

2. Nearly 22,000 multiple-choice items on tests given to college classes were canvassed for such changes. Of the 555 changes clearly distinguished, fifty-three per cent raised the score, twenty-one per cent lowered it and twenty-six per cent had no effect on the score.

3. Nearly 18,000 true-false test items were likewise examined for changed responses. Of 570 changes tabulated, those tending to raise the score constituted sixty-three per cent. Thirty-four per cent tended to lower the score and three per cent had no effect on it.

4. A comparison of the changes made by thirteen superior and thirteen inferior students showed that the inferior students made more changes. Only forty-nine per cent of their changes tended to raise their scores, however, while sixty-eight per cent of the changes made by superior students tended to raise their scores.

5. Students should be informed that it pays in terms of score to check over all questionable items carefully in true-false and multiple-choice types of tests rather than to trust first impressions. They may expect to raise their scores at least twice as often as they lower them by changing their first responses when later judgment seems to justify it.

THE INFLUENCE OF AFFECTIVE TONE ON LEARNING AND RETENTION

RUTH M. CHANEY

Iowa State College

AND

ALVHH R. LAUER

Ohio State University

Statement of the Problem

Much has been said concerning the effect of different types of subject-matter on retention, and usually certain emotional elements are assumed to be basic to learning.

The law of effect must first be considered in this connection and as stated by Gates,[1] "The individual tends to repeat and learn quickly those reactions which are accompanied or followed by a satisfying state of affairs." In man it is assumed that a satisfying state of affairs is accompanied by the feeling of pleasantness; while an annoying state of affairs is accompanied by unpleasantness. Thorndike[3] defines the above law: "When a modifiable connection between a situation and a response is made and is accompanied or followed by a satisfying state of affairs, its strength is increased. When made and accompanied or followed by an annoying state of affairs, its strength is decreased." This is slightly different from Gates' interpretation of Thorndike and is somewhat less inclusive. It seems reasonable to believe that subtle changes occuring in the neurone under certain conditions would modify the effect of responses accompanied by these states. Therefore, logically it would be expected that material which is interesting would be easier to learn or memorize and retain than material which is uninteresting. Likewise agreeable material would be more easily learned than that which is disagreeable. Pleasantness would thus by its very nature and *per se* cause the neurone to be easily modified and habits would be more readily formed when accompanied by a satisfying state of affairs.

Theoretically, we should have here a basis for efficient teaching. To make one's teaching highly effective, subject-matter that is pleasant to the student should be used. Likewise any type of punishment would be justified since reactions which are accompanied or

followed by unpleasantness tend not to be repeated. Practically these postulates need to be adequately verified by actual experiment rather than by common observation which many times is misleading. This need is enhanced by the fact that the whole problem of motivation hinges on the assumption laid down by Thorndike in his satisfiers and annoyers.

Smith[2] found that certain words which had greater affective tone were best remembered. This was true irrespective of whether it was negative or positive. In other words the impression made upon the learner seemed to have more to do with retention than the fact that the stimulus was pleasant. It might be either satisfying or annoying.

Our purpose was to present material of graded degrees of pleasantness or its opposite under controlled conditions and to measure the amount of retention after a given number of repetitions; attitude, purposive and attention factors being kept fairly constant.

Method and Procedure

Twenty girls ranging in age from eighteen to twenty-one were used as subjects in this experiment. These girls were students at Iowa State College and were for the most part freshmen and sophomores. They were divided into four groups of five and accordingly four different lists of words were used. The first, list (A) consisted of pleasant words; the second (B) neutral words; the third (C) unpleasant words; and the fourth (D) words selected from the three previous groups. There were twelve words in each list. The words had been evaluated by a group of one hundred fifty students on a five point scale and were weighted according to these values. Hence the words were approximately equal in affective tone if such could be measured in this way. A fifth list (E) of nonsense syllables was used as a control. Each group was given a different list of words, A, B, C, or D, but all were given the list E. In this way group differences were thought to be equated.

The test was given after lunch during the noon hour for five consecutive days. On the first day each subject was given thirty seconds to study her list of twelve words. Then she was asked to write the words recalled. Next a list of nonsense syllables was given and the same procedure followed as for the previous test of meaning words. For the following three days the subjects were merely asked to study the words for thirty seconds. On the fifth day the subjects were again

asked to write the words and nonsense syllables. The two lists were alternated, that is on the first day the meaning words were given first; on the second day, the nonsense syllables first; and on the third day the meaning words first, etc. The subjects were not told in advance whether they were to write or not. To offset the effect of primacy or serial position the word order was changed for each sitting.

Results

The mean number of words for each group was computed both on the meaning and nonsense lists. In order to compare the groups, the results were expressed in a ratio of the mean of sense words to the mean of nonsense syllables recalled by each. This may be called the coefficient of coincidence, although it would be expected to be greater than unity since meaning words should be remembered better than meaningless syllables.

Table I

Group	Words	Per cent meaning words recalled	Per cent nonsense syllables recalled	Ratio
A	Pleasant words	72.2	61.6	1.172
B	Neutral words	71.4	60.5	1.180
C	Unpleasant words	72.9	62.5	1.166
D	Mixed words	80.6	71.0	1.135
	Mean	74.5	63.9	1.163

Conclusions

1. Since the probable error of the expressed ratios is approximately $\pm .023$, and the greatest deviation is scarcely larger than one PE, it would seem that there is no significant difference in the learning of such materials as were used in this study.

2. With respect to the advantage of meaningful material over nonsense material in retention (a secondary point brought out in the study) it was found that 16.3 per cent more meaningful material was recalled.

3. The results seem to indicate a certain common error of logic in attributing the success of learning "under pleasant conditions" or

"learning interesting material" to some mythical force called "effect" which is usually considered as an aspect of feeling tone.

4. Whether or not the two above described conditions may be synonymous, the "law of effect" would appear to be merely the attitude of the learner which could be easiest explained as intensity of association during the learning process.

5. Although the number of cases was not large enough to warrant general conclusions, it would seem from our data that factors other than mere affective tone would be highly potent in facilitating learning of the type used in this experiment if such facilitation was in evidence.

References

1. Gates, Arthur I.: "Psychology for Students of Education." Macmillan Co., 1926, pp. 230–235.
2. Smith, Whately W.: "Measurement of Emotion." Harcourt, Brace Co., 1922, pp. 29–56.
3. Thorndike, E. L.: "Educational Psychology." Teachers College, Columbia University, N. Y., 1923, p. 172.

TRANSFER OF TRAINING IN THE CANCELLATION EXPERIMENT

B. H. NEWKIRK AND R. H. GUNDLACH

University of Washington

In a recent book Poffenberger draws certain conclusions about the transfer of training which, it appears to us, are not justified by his experimental findings.[1] We have repeated his experiment with a few essential modifications in procedure, with results which seem to justify our criticisms.

In Poffenberger's experiment eight subjects repeated five times the cancellation of groups containing 3's and 5's, and the cancellation of groups containing 4's and 7's. Four of these subjects then became a training group, while the others became the control group. The training group was practised in the cancellation of 3's and 5's separately. This training continued for nine days with approximately eleven tests each day. On the tenth day the combined group of trained and control subjects repeated the tests used at the beginning of the experiment. Five trials were given as at the start. The first two trials, both of the initial and final test series, were not counted in the results. Poffenberger estimated the improvement from separate cancellation of 3's and 5's to the cancellation of groups containing 3's and 5's to be about eight per cent. There was no improvement in the cancellation of groups containing 4's and 7's.

Poffenberger makes this statement, "In the one case we find identical elements, the numbers 3 and 5, responsible for the transferred improvement, and in the case of the numbers 4 and 7 there were no identical elements, and no improvement. This means that there was no improvement in the general ability to cancel numbers, but only in the ability to cancel specific numbers or groups of numbers."[2]

Several objections to Poffenberger's procedure arise. In the test series there was no transfer of training from 3's and 5's to 4's and 7's probably because there would be interference in cancelling 4's and 7's, both from the negative adaptation of the subjects to those numbers and from the serial habit established in following the spatial arrange-

[1] Poffenberger, A. T.: "Applied Psychology." Appleton and Co., New York, 1927, p. 75. His evidence comes from an earlier experiment, The Influence of Improvement in One Simple Mental Process upon Other Related Processes. *Jour. of Educational Psychology*, Vol. VI, 1915, pp. 459–474.

[2] "Applied Psychology." P. 75.

ment of 3's and 5's on the practise blank. The training subjects practised the cancellation of 3's and 5's on a single form, so that they were not only learning to note and cancel these numbers and to ignore other numbers, but were also learning a certain spatial arrangement of the 3's and 5's on the blank.

Other less important objections to Poffenberger's procedure may be listed. (1) There were but eight subjects, with but four in the training group. (2) Poffenberger's subjects were chosen from a partly trained group, familiar with the material used and (3) familiar with the results which Poffenberger wished to obtain. It is possible that the knowledge of when transfer was not expected might have affected the performance of the subjects in the control group.

To test out these criticisms we modified Poffenberger's procedure in at least three respects: (1) In order to eliminate the tested items from the practise series, the transfer of training was measured from cancellation of numbers to letters, (2) to elimate the formation of position habits, different forms were used in the training and test series; (3) to increase the reliability a large number of subjects was used. The two hundred fifty subjects were made up of the introductory class in psychology during the winter quarter of 1928. None knew the purpose of the experiment.

The items of the different cancellation blanks were arranged in a chance order. On every printed blank, 8½ inches by 11 inches, there were twenty lines of fifty figures or letters each. Seven different forms of the figure blanks were prepared, and two different forms of the letter blanks.

The subjects were given no other directions than to cancel, across the rows, all the 6's (or D's) they could find. When the training group was practising the control group stood up with eyes off the work of the training group, and timed the training subjects. In correcting the blanks, errors were compensated by Poffenberger's method of adding to the actual time a correction for mistakes—the number of errors multiplied by the average cancellation time (for each cancelled item).

The results of transference were made more comparable by concluding the final test series with the cancellation of another number, 2. Here, as in Poffenberger's experiment, the subject had to face and ignore the very item (6) he had been trained to cancel. If this situation has little significance, then the practised group should do relatively as well on the 2's as on the D's.

The learning curve for the practise group (Table I), although it shows considerable rise, has not flattened out. The initial and final test results are shown in Table II.

TABLE I.—RESULTS OF THE TRAINING GROUP ON THE CANCELLATION OF 6's

Week	Mean	SD	SD_m
First week	171.55	35.89	3.39
Second week	143.63	25.31	2.39
	128.74	20.63	1.95
	128.09	19.31	1.83
Third week	119.4	19.87	1.88
	112.85	18.32	1.73
	110.88	19.37	1.83
	117.2	20.49	1.93
Fourth week	109.45	23.39	2.21
	103.62	17.15	1.62
	104.65	17.92	1.69
	107.9	18.11	1.71
Fifth week	100.9	17.95	1.7
	99.61	19.35	1.82
Sixth week	103.82	17.03	1.6

Despite the inferiority of the training group on the initial trials they show a marked superiority over the control group in the final cancellation of D's. Such a difference would occur in unrelated material about four times in one thousand cases. The actual gain in seconds of the final over the initial trial on the letters has been averaged. A difference between the means of two groups as large as this, and in this direction, would occur by chance less than once in 1,000,000 times. The improvement of the practised group is about 7.5 per cent more than the improvement of the control group.

Inspection of the data from the cancellation of 2's shows a different result. The experiment here is closely analogous to Poffenberger's in that the effects of practise in cancelling one number are measured on the cancellation of another number. Our results in this agree with his in showing that the practised group has gained practically nothing from its training.

In summary we find, concordant with Poffenberger's results, but contrary to his conclusions, that practise in the cancellation of one item aids in the speed of cancelling another.

TABLE II.—COMPARISON OF THE INITIAL AND FINAL RESULTS OF THE TRAINED AND CONTROL GROUPS

Measures		A, training group (112)	B, control group (138)	Difference, A − B, $SD_{diff.}$
Initial cancellation of 6's	Mean	171.55	167.19	
	SD	35.89	27.05	1.06
	SD_m	3.39	2.3	
Final cancellation of 6's	Mean	103.82	133.94	
	SD	17.03	20.17	−12.87
	SD_m	1.6	1.71	
Initial cancellation of D's	Mean	147.65	142.28	
	SD	24.95	20.48	1.83
	SD_m	2.36	1.74	
Final cancellation of D's	Mean	111.26	119.92	
	SD	29.66	19.37	− 2.66
	SD_m	2.8	1.65	
Cancellation of 2's	Mean	142.46	140.0	
	SD	25.52	19.0	.84
	SD_m	2.41	1.62	
Initial—final time, cancellation of D's.	Mean	36.63	24.56	
	SD	19.81	18.81	− 4.91
	SD_m	1.87	1.6	

THE CONSTANCY OF IMAGINAL CONTENT

HENRY BOWERS

Fergus High School, Ontario

The writer has been carrying out experiments on factors influencing the retention of words isolated from context. In this connection the clarity of the subject's visual or auditory image of the object or sound denoted by the word came under investigation.

Visual Words

At the beginning of a period classes in the Fergus High School were addressed thus:

"In our thinking some of us frequently see mental pictures or images, as they are called. How many of you can form a mental picture of a buffalo? (Practically every one held up his hand.) Can you see its head clearly? Its hoofs? Its tail? Do you think that all of you see that buffalo with precisely the same degree of clearness? . . . Well, to-day I should like to see how you form your mental pictures. This is not an examination. There will be nothing to count, for or against you. Write these words down in a column—diamond, lamp, red, star, stain, flag, spark, was, tar, picture, sunset, house, whisker, shine, blue, deep, inch, valley, elephant, cow, alas, curly, three, forest, fog, desert, snow, ocean, that, muddy, twinkle, horse, car, spider, suds, tombstone, green, skeleton, spear, darky, flame, yellow, dusk, blot, plate, rat, mountain, black, tree, footprint, handsome, blaze, foam, moon, tower, sword, tarnish, swan, cloud, glitter, angel, pink, yet.

"In front of each word where you have a clear distinct mental picture of the object referred to, write the number 3. If your image is only fairly good, rather vague and indistinct, write the number 2; if quite poor, very vague and indistinct, write 1. If you have no image of the object write 0."

The following was written on the blackboard:

3. If your mental picture is very good, quite clear and distinct.
2. If your mental picture is only fairly good, rather vague and indistinct.
1. If your mental picture is very poor, quite vague and indistinct.
0. If you can form no mental picture of the object.

When finished, go over your list a second time reconsidering each claim.

These "detectives" rubbed shoulders with the crowd—alas, that, was, yet. It was felt that subjects claiming visual images in these four cases should be ruled out as an unreliable.

At the end of a week the list was given to as many of the same pupils as were present. The images for each word claimed for each pupil were recorded. Those who did not record "0" in front of each of the detective words both times were eliminated. Also, whenever opposite any word the same pupil recorded two numbers differing by more than "1," that pupil's data for the words were discarded.

Rather more than fifty per cent of the subjects were classified as "unreliable," *i.e.*, they claimed images varying from "poor" to "very good" for one or more of the "objects" denoted by "alas," "was," "yet," on at least one of the two trials. Senior forms were equally culpable with junior ones in this respect. Only five per cent of words were rejected for having grossly disparate images assigned by the same pupil. In tabulating the data the sexes were kept separate. Each alternate pupil, from an alphabetical arrangement, was selected, and the mean image for twenty words was found. Also, the mean image of the remaining subjects was found. There were thirty-one reagents in each class.

Pearson's r was calculated between the mean claimed image for each word. It was $.86 \pm .04$. To find if there was any sex-differences the mean image for each word was found for all boys and also for all girls. Pearson's R for these two sets of values was .64, giving $r = .86 \pm .03$.

To find if the "unreliables" really did constitute a separate class, the coefficient of correlation between the mean claimed "reliable" and the mean claimed "unreliable" images was calculated. Pearson's r was $.79 \pm .05$. Clearly there is little or nothing to be gained by exclusion of the so-called unreliable reagents. When dealing with "auditory words" the trap was not set.

Mean reliable images (decimal point omitted) are given below. The first number is that obtained from the boys (thirty-six subjects), the second from the girls (twenty-five subjects):

horse (297, 300), house (296, 296), lamp (290, 296), flag (290, 288), snow (290, 296), cow (289, 292), car (286, 300), forest (285, 290), darkey (281, 273), plate (281, 296), tree (279, 300), moon (278, 298), star (276, 284), elephant (271, 267), cloud (266, 283), mountain (270, 261), spider (270, 282), tombstone (270, 292), picture (269, 285), sunset (267, 294), black (266, 275), sword (265, 231), tower (263, 256), rat (262, 269), valley (258, 260), footprint (257, 275), blaze (257, 258),

flame (253, 277), swan (253, 246), foam (251, 220), red (249, 263), spear (245, 212), ocean (244, 258), blue (243, 252), tar (241, 256), muddy (240, 254), diamond (239, 273), blot (239, 275, spark (238, 252), green (234, 260), fog (233, 252), suds (230, 256), curly (227, 238), whisker (226, 269), skeleton (226, 234), yellow (223, 270), stain (221, 260), inch (212, 248), pink (207, 250), desert (203, 204), dusk (198, 204), angel (190, 235), three (179, 238), shine (175, 196), deep (161, 198), twinkle (147, 159), handsome (143, 150), glitter (138, 170), tarnish (108, 159).

Auditory Words

The pupils were addressed as follows: "I am very much interested in the way in which we hear various sounds in our mind's ear. For, just as we have a mind's eye, as you may remember in connection with our mental pictures, so we have also a mind's ear. Can you hear in your mind's ear the noise made when tearing a piece of cloth? Some of you may have a very clear image of the sound, others only fair. Some may have no image at all. Write down the following words in column: hiss, bang, slam, growl, crackle, yell, babble, caw, mumble, chirp, blare, grunt, quack, yap, splash, groan, whine, thud, mew, moan, click, bleat, hum, purr, scream, call, shrill, rattle, cry, tinkle, gurgle, moo, thunder, squeal, hallo, gabble, clang, hoot, cackle, yelp, pop, murmur, boom, roar, bellow, crunch, clap, toot, mutter, ha, whimper, cluck, ding-dong, jingle, shriek, scold, chime, coo, shout."

The following was written on the black-board:

Write after each word the appropriate number.
3. If you can hear the sound very clearly and distinctly in your mind's ear.
2. If you hear the sound only fairly well, not very clearly.
1. If you can hear the sound poorly, very ill-defined and indistinct.
0. If you cannot hear the sound at all.
This is a test for the words, not of you. When finished go over the list again.

The image claimed for each word was recorded for each pupil. In view of the fact that serious differences in the claimed visual images for a given word for a given pupil existed in only five per cent of the cases, it was not considered necessary to repeat the test. Also, the established ability of the "unreliables" to sort out visual words in largely the same order as did the "reliables" led me to neglect such a distinction in this case.

The twenty boys and twenty girls of Form 1 were arranged alphabetically, the sexes being kept separate. Pearson's r for the mean claimed images for twenty auditory words for alternate words was

.71 ± .02. This was also calculated for the mean image claimed by boys for sixty sounds and that claimed by the girls. It was .75 ± .04.

Mean images (decimal point omitted) are given below. The first number is that obtained from the boys (twenty subjects), the second from the girls (twenty subjects):

hiss (230, 250), bang (280, 290), slam (260, 270), echo (255, 240), growl (275, 270), crackle (230, 220), yell (270, 260), babble (225, 180), caw (265, 255), mumble (220, 220), chirp (245, 250), blare (53, 80), grunt (260, 255), quack (270, 200), yap (205, 170), splash (285, 285), groan (255, 244), whine (274, 258), thud (250, 225), mew (245, 290), yelp (280, 275), pop (240, 205), murmur (245, 225), boom (260, 205), roar (285, 290), bellow (220, 185), crunch (225, 205), clap (295, 285), toot (270, 255), mutter (230, 220), ha (210, 215), whimper (225, 240), cluck (260, 215), ding-dong (280, 260), jingle (285, 240), shriek (250, 250), scold (235, 245), chime (250, 275), coo (235, 245), shout (280, 290), moan (260, 255), click (265, 245), bleat (190, 145), hum (243, 275), purr (285, 280), scream (270, 285), call (255, 250), shrill (190, 185), rattle (245, 230), cry (247, 250), tinkle (250, 240), gurgle (225, 220), moo (235, 252), thunder (270, 285), squeal (270, 250), hallo (265, 245), gabble (205, 155), clang (250, 245), hoot (220, 180), cackle (260, 263).

Conclusions

1. In a given group there is a highly marked tendency to establish a constant order of precedence in clarity of visual and auditory images; or, there is a marked tendency for the object or sound represented by a particular word to have a constant imaginal content in a particular group of reagents.

2. There is little or no sex-difference in the order in which visual and auditory images are arranged on the basis of clarity.

SOME RESULTS OF THE MULTI-MENTAL TEST

PAUL A. WITTY AND J. F. TAYLOR

University of Kansas

The problem of securing reliable data in reference to the mental ability of school children is a difficult one for the administrator. The Stanford Revision of the Binet Test of Intelligence cannot be used by many schools because it is too costly from an administrative standpoint since its use posits specially trained examiners and individual time-consuming observations. The group intelligence test has been proposed as an efficient and practical substitute. However, some of the group mental tests yield relatively unreliable measures of mental ability. It is possible, of course, to secure fairly accurate data for use in classifying pupils by combining the results of two of the better group tests.[1] Such a procedure is however, costly and in many school systems impossible.

There is need for a single, short, reliable group test which can be used as a substitute for the Binet individual examination. Professor McCall has attempted to provide such a device.[2] His Multi-mental Scale is a short group test which is simple to give and score. One test and one sheet only is needed for examining each pupil; therefore the cost is decidedly low.

The vailidity of such a test is determined by the extent to which it accomplishes in practical use the ends for which it is devised. Obviously, one criterion that may be used to determine the worth of a mental test is the degree to which it yields results comparable to those obtained by employing the Stanford Revision of the Binet Test. Another criterion of the value of a mental test is found in the extent to which its results correlate with academic success. This latter item is of particular moment if the test is to be used for classifying pupils in school.

The following paper deals with an evaluation of the Multi-mental Test. The writers have assembled data showing: (*a*) the correlation between the results of the Multi-mental and Stanford-Binet Tests for unselected elementary school pupils of Grades IV, V, and VI; (*b*) the

[1] See Brooks, Fowler D.: The Accuracy of Intelligence from Pairs of Group Tests in the Senior High School. *Journal of Educational Psychology*, March, 1927, pp. 173–187.

[2] *The Teachers College Record*, Oct., 1925 and Jan., 1926, pp. 109–120, 394–415.

correlation between the results of the Stanford Achievement Test and the Multi-mental and Binet Tests respectively for the same children.

Data were secured for five hundred twenty-two Lawrence, Kansas, school children distributed according to grade as follows:

	Number
Fourth	88
Fifth	178
Sixth	256
Total	522

The Multi-mental Tests were given to the children the last week in October, 1926. At the same time the Stanford Achievement Test, Form A was administered. The Stanford Revision of the Binet-Simon Test of Intelligence had been given to these children during the years 1923–1926 inclusive. The results of this test were assembled. The intelligence quotient was assumed to be relatively constant and each child's mental age on October 28 was estimated from his IQ previously secured. The three ages, mental age upon the Binet Test, mental age upon the Multi-mental, and educational age thus were made comparable.

The Pearson Product Moment Formula was employed to disclose the relationship between the three ages. Table I displays the coefficients of correlation.

Table I.—Coefficients of Correlation (Pearson Product Moment Formula)

r between Binet MA and Multi-mental age	$= +.534$
	PE $= .021$
r between Binet MA and Stanford EA	$= +.634$
	PE $= .017$
r between Multi-mental age and Stanford EA	$= +.78$
	PE $= .012$

It is interesting to note that the Multi-mental Scale does *not* yield results sufficiently comparable to those secured from the Binet Test to warrant its use as a substitute for the Binet examination. This fact becomes more conspicuous when the data are partitioned according to grade. These data are presented in Table II.

Analysis of the Multi-mental Test shows it to be essentially different from the Binet Test in content. Intelligence was interpreted by McCall as general intellectual ability and the emphasis in the construc-

tion of the test was placed upon the intelligence factor essential in school work. The Stanford-Binet Test was devised with a broader concept of intelligence in mind. It is not surprising therefore that the two tests show no closer agreement. In constructing the Multi-mental Scale the criterion score with which the results of the test were compared was made up of the following elements: the teacher's judgment, Stanford-Binet, Stanford Achievement, National Intelligence, Thorndike-McCall Reading, Morrison-McCall Spelling, Woody-McCall Mixed Fundamentals in Arithmetic. Of the agreement of the mental test with this criterion, the makers of the Multi-mental Scale state:

TABLE II.—COEFFICIENTS OF CORRELATION (PEARSON PRODUCT MOMENT FORMULA)

Grade IV r between Multi-mental and Binet age	$= +.30$
PE	$= \pm .046$
Grade V r between Multi-mental and Binet age	$= +.306$
PE	$= \pm .045$
Grade VI r between Multi-mental and Binet age	$= +.275$
PE	$= +.065$

"The Multi-mental Scale was given to a group of ninety pupils ranging from the third through the eighth grade who had not been tested with the Multi-mental Scale before. A preliminary comparison of the correlation in this study shows that the Multi-mental Scale correlates higher than either the Stanford Revision of the Binet-Simon Intelligence Scale or the National Intelligence Test with the Criterion Score, namely, the composite of a dozen tests including the Stanford-Binet Scale and the National Intelligence Test. The correlation of the Stanford-Binet Scale and the National Intelligence Test with the Criterion Score is .89 and .93, respectively. The correlation of the Multi-mental Scale with the Criterion Score is .93."[1]

McCall concludes that one form of the Multi-mental Scale is more valid than the Binet Scale and almost as valid as one form of the National Intelligence Test, for measuring the ability indicated by the Criterion cited above. He states further that since two forms of the Multi-mental Test can be given and scored in the time required for one form of the National Intelligence Test, the Multi-mental Test represents a forward step in test construction.

It is indeed interesting that the Multi-mental Test gives a higher correlation with the Stanford Achievement Test than does the Stan-

[1] *Teachers College Record*, January, 1926, p. 413.

ford-Binet Scale (see Table I). The value of a test is always relative. If a test of intelligence is to be used chiefly in school, one important criterion of worth is the degree to which it correlates with an adequate criterion of school achievement. The Stanford Achievement Test is an excellent educational achievement assembly, covering in a comprehensive manner the various items of learning stressed in the elementary school. This test has a high degree of validity and reliability. Since the Multi-mental Test gives a high coefficient of correlation with the Stanford Achievement Test, it seems reasonable to assume that this mental test will prove a decided asset to the administrator who seeks a short, reliable testing tool to be used in classification of pupils and in prognosis of their educational achievement.

PERFORMANCE TESTS AND FOREIGN CHILDREN

OLIVE P. LESTER

Psychology Department, University of Buffalo

Many psychologists and educators have felt the inadequacy of the Binet, Pressey, Otis and other similar verbal tests in measuring satisfactorily the intelligence of foreign children (this applies more particularly to those foreign children who enter the primary grades where as yet they have had very little opportunity to learn English) and as an outgrowth of this, there has resulted an attempt to use these standard tests after translating them into different foreign languages. In instances where this method has been tried the IQ ratings or mental ages of the foreign children are found to be higher.[1]

Another method of approach to the problem of testing and classifying the foreign child more adequately has been through the performance tests. Here the handicap of language is reduced if not completely eliminated. This study is concerned with the results of administering such tests to a group of foreign children, and with the possibility of using performance tests more widely in the sectioning of foreign children in our school systems.

The main group studied was a retarded first grade in one of the public schools of the city composed of twenty-six children, only two of whom had kindergarten training. The classification was based on the Pressey Primary Test. This seems to be quite a prevalent practice—the administration to foreign children of a group test dependent largely on language ability and the subsequent sectioning on this basis. As eighteen of these children were from homes where Polish was spoken entirely, the question arose as to how far the classification "retarded" was due to actual lack of ability and how far to language handicap, and what would be the result of giving performance tests in such a group. On the whole this group seems fairly representative of most first grades in a foreign section of a large city. The performance tests were also administered to the bright division of the same first grade and results noted.

---

[1] Brown, G. L.: Intelligence as Related to Nationality. *Journal of Educational Research*, Vol. IV; An Experiment with the Otis Group Intelligence Scale. *School Magazine*, Buffalo, Vol. IV, No. 10, June, 1922.

The tests employed were the Seguin Form Board, Mare and Foal Test, Knox Cube Test,[1] Healy A Form Board, Kohs Block Test[2] and the Porteus Maze. The Mazes standardized for ages five, six, and seven in both the foreign group and the bright group gave very unsatisfactory results, the experimenter apparently being at fault in giving directions; but the instructions for understanding the problem at hand are rather involved and seem incomprehensible to a foreign group at least when used as a group test.

The Seguin Form Board was found to correlate with chronological age more than with mental age—success in it seems to be based to a large extent on physical development and muscular coordination; hence this test has been eliminated in the discussion of the results.

TABLE I.—SHOWING CORRELATIONS BETWEEN BINET MENTAL AGE AND PERFORMANCE TESTS

	Number of cases	Mare and Foal	Kohs Block Test	Healy A Form Board	Knox Cube Test	Combined tests (5) including Seguin	Selected tests (Kohs and Mare and Foal)
Binet mental age....	26	.65	.60	.46	.28	.64	.66

From Table I it is seen that certain of the performance tests, that is, the Mare and Foal, and Kohs have a good deal in common with the Binet Test. In the bright division the Mare and Foal and Kohs also show the highest correlations with mental age, +.25 and +.40 respectively. Apparently the language factor involved in giving directions for these performance tests and the problems involved in the tests necessitate an acquaintance with a rather typical American environment.[3] In other words we might classify these performance tests more as "culture tests."

The Healy A Form Board and the Knox Cube Test which require in their administration practically no language show fairly significant correlations with mental age. It is also in these two tests, as will

[1] Eight combinations or lines were used, five of these were identical with Pintner's; one differed only in one number and two lines were entirely different.

[2] Kohs' directions were followed except in scoring and there only slight deviations were made.

[3] Both the Mare and Foal Test and the Kohs correlated fairly high with chronological age, .31 and .52 respectively—a longer acquaintance with an environment resembling that of an American child apparently proves advantageous for success in the tests.

be shown later, that the Polish children actually demonstrate their true ability divorced from language handicaps. When correlated with teachers' estimate, the Healy Form Board and the Knox Cube Test show a positive correlation of .35 and .54 respectively, while the other two performance tests show correlations of −.05 in case of the Mare and Foal and +.28 in the Kohs.

TABLE II.—SHOWING THE CORRELATIONS BETWEEN TEACHER'S ESTIMATE AND THE PERFORMANCE TESTS, AND THE PRESSEY TEST

	Mare and Foal	Kohs Block Test	Healy A Form Board	Knox Cube Test	Pressey Test
Teacher's estimate.....	−.05	.28	.35	.54	.31

The teacher in estimating the intelligence of the children was asked to include under this heading mainly the kind of school work done by each child, his abilities and capacities for leadership as manifested in play, and his general alertness, understanding and interest in the "school situation." A scale rating from 1–5 was used; 1, being very inferior; 2, inferior; 3, average; 4, superior; 5, very superior. The question will undoubtedly be raised as to the weight to be attached to teachers' estimates. By some, this criterion may be classed as valueless, but it seems that at the present time at least this is one of the best means we have of judging tests. In a class of foreign children it seems plausible that a teacher's estimate of the child's capacity (taken as it is of the "entire" child in the classroom situation) is more adequate than a verbal test which places the foreign child at such a disadvantage. Example after example might be cited of the checking up and evaluating of the various kinds of tests by teachers' estimates concerning the individuals tested. Cyril Burt adopts as his criterion for testing the tests the judgments of the class teacher or principal and in justifying the procedure states: "There is no standard of comparison which can surpass or supercede the considerate estimate of an observant teacher working daily with the individual children over a period of several months or years."[1]

[1] Burt, Cyril: "Mental and Scholastic Tests." P. 199.

The teacher of this foreign group had been associated with these children for a period of seven months before her opinion as to their general intelligence had been asked for. This young woman was not only a college graduate, but had specialized in educational work, testing, etc. One interesting check is available on her estimates of intelligence. She knew the IQ ratings obtained by the Pressey tests, but did not know the ratings found when the Binet test was used. The correlation however, between her estimate and the Pressey ratings was +.31, while with the Binet ratings it was +.44.

Apparently then, if we concede any value to the teacher's estimate, the Healy and Knox tests give a better estimate of a child's ability than do either of the other two performance tests. The Knox test alone could be given to a foreign group of children and results obtained which correlate better with teacher's estimate than do Binet tests—correlations being +.54 and +.44 respectively; and in this test, language ability is practically uncalled for. This finding is not at all applicable in a group of English speaking children.

Examining the scores within the group on the various performance tests bring out some interesting results. As was noted above eighteen of the twenty-six children came from homes where Polish was used entirely. The remaining eight, or nearly one-third of the grade were American children with the exception of one Polish boy. As his family use English in the home, he has also been classified in the English-speaking group. The following table shows the median scores obtained in these tests by the English and non-English speaking groups:

TABLE III.—SHOWING MEDIAN SCORES IN TESTS ON THE BASIS OF LANGUAGE SPOKEN IN THE HOME

	Pressey	Binet IQ	Mare and Foal	Kohs Block Test	Healy A Form Board	Knox Cube Test
English speaking group	81.5	81	5′ 18.5″	6.25	3′ 23.5″	2.5
Polish speaking group..	69	80.5	6′ 43.7″	8	1′ 48.8″	5.25

In the Pressey test the Polish group stands out as definitely inferior; in the Binet test both groups scored essentially the same, but in these tests the Polish group was decidedly handicapped by language factors. In the Mare and Foal test the English speaking

group is again superior, but in the Kohs, and Knox and Healy, though particularly in the latter two tests, the Polish-speaking children stand out as being decidedly superior. It seems that here we have a case of foreign children under language handicaps testing the same or lower than a section of English-speaking children and being grouped with them in our school system. Yet these American children are definitely inferior in intelligence. Here is an instance where a grave injustice is done the foreign child. Undoubtedly his lack of facility with the English language makes it impossible to put him in with American children who are his equal mentally; but is the solution of the difficulty the classifying him with American children who are dull or borderline? The answer seems to be in the negative. Classes in English must be organized for him, and when he has mastered the rudiments of language, to place him back with American children of his own intelligence seems to be the only wise and just procedure.

Now in conclusion, let us turn for the moment to the practical problem of how more adequately to test the abilities of the foreign child for the purposes of classifying him in a grade. Again using teachers' estimate as a criterion we find that this can be accomplished by combining certain of the performance tests with the Binet test.

TABLE IV.—SHOWING CORRELATIONS BETWEEN TEACHERS' ESTIMATES AND BINET AND PERFORMANCE TESTS

	IQ ratings	Combined performance tests (5)	Binet and combined tests (5)	Combination of Binet, Healy and Knox tests
Teachers' estimates (foreign group).....	.44	.35	.41	.56
Teachers' estimates (bright division first grade).....	.46	.16	.50	.51

It seems reasonable to conclude that, in combining the Healy and Knox tests with the Binet, we have a better means of classification for foreign children than that given by the Binet test alone, or by the combination of the Binet test and unselected performance tests, since some of the tests present the same difficulties to the foreign child as are met with in the Binet. Furthermore, such a combination of Binet, Healy and Knox is somewhat advantageous in testing an American group of children as the above table shows.

It must be realized that no very adequate generalizations can be based upon such a small number of cases, yet perhaps the material here obtained may prove suggestive for further work along similar lines.

Since a check was desired upon the results of combining teachers' estimates and performance tests with the Binet test, the Healy A Form Board and the Knox Cube test were given this year to two first grade classes of Polish children who had already been given the Binet test. The teachers of these grades were asked to rate the children as to their intelligence. The same method of scoring was used by them as had been employed previously by the other two teachers. One of the groups of twenty-three children, Group A, let us say, was almost identical with the one previously studied. It was the most retarded section with a median IQ of 76 and contained four American children: the other, Group B, a group of all Polish children, twenty in number. This latter group was classified as average in the school system and had a median IQ of 93. Incidentally it is interesting to note that these Polish children who were in the average division of the first grade, on the whole, came from homes where English was spoken, and they themselves could use the language more fluently.

Table V seems to indicate the effect of facility in the use of language. In the Binet test where verbal ability is important Group A is definitely inferior to Group B, while in the performance tests their scores are comparable.

TABLE V.—SHOWING THE MEDIAN SCORES OBTAINED BY THE POLISH CHILDREN IN GROUP A AND THE SCORES OBTAINED BY GROUP B

	Number of cases	Binet Test	Healy A Form Board	Knox Cube Test
Group A.....	19	74	10′ 9.1″	5
Group B.....	20	93	10′ 10″	5.25

TABLE VI.—SHOWING CORRELATIONS BETWEEN TEACHERS' ESTIMATES AND THE BINET, HEALY, AND KNOX TESTS, AND COMBINATIONS OF THESE

	Binet Test	Healy A Form Board	Knox Cube Test	Binet, Healy and Knox
Teacher's estimate of Group A.....	.05	.36	.17	.31
Teacher's estimate of Group B.....	.09	.21	.20	.30

The correlations between teachers' estimates and the Binet test alone, are much lower than in the study reported, yet one does find that the correlations increase rather significantly when a combination of Binet, Healy and Knox is compared with Binet alone.

Conclusions

1. Certain performance tests apparently resemble the Binet test more closely than do others (in our study the Mare and Foal and the Kohs Block test), and in working in groups of foreign children this must be kept in mind.

2. Some of these performance tests, the Healy Form Board and the Knox Cube test, were found in the group studied to correlate better with teachers' estimates than did others.

3. Within the group after it was divided on the basis of language spoken in the home, the Polish children proved themselves definitely superior in the Healy and Knox tests when they were not handicapped by lengthy verbal instructions. It appears in this group at least that the foreign children have more ability and ought not to have been grouped with dull and borderline American children.

4. In classifying the children of foreign parents it would seem more practical to combine performance tests with Binet tests, though here it would be more desirable to combine selected performance tests (those least resembling the Binet test).

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY JAMES E. MENDENHALL

Research Associate, Institute of Educational Research, Teachers College
Columbia University

Intelligence Testing

Jews and Others. Some Group Differences in Personality, Intelligence, and College Achievement. Henry E. Garrett. The Personnel Journal, Feb., 1929, 341–348. Within a sample of two hundred ninety-six Columbia College freshmen, the correlation between college grades and scores on the Thorndike Intelligence Examination was .63; between Thorndike and Social Intelligence Test, .57; between Social Test and Grades, .38; Personal Inventory (Colgate Mental Hygiene Tests) with Grades, .11, and with Thorndike, .05, and with Social Intelligence, .03. Jewish students rank significantly higher than other freshman groups classified as to national origin both in grades and Thorndike score.

Need of Equating Intelligence Quotients Obtained from Group Tests. Grayson N. Kefauver. Journal of Educational Research, Feb., 1929, 92–101. The writer presents a table of comparable IQ's for twelve tests. "The variation in the means of the intelligence-quotients on different tests for a group of pupils was twenty-three points."

Analysis of Specific Responses of Children in the Healy Pictorial Completion Test II. Mildren Day Dorcus. Journal of Genetic Psychology, Dec., 1928, 574–586.

Group Tests for Clinical Studies. Grace H. Kent and David Shakow. Journal of Genetic Psychology, Dec., 1928, 595–618. A description is given of procedure and of tests especially suited to clinic work. Tests of information, similarity, essential difference, arithmetical reasoning, Holley vocabulary, and modified Trabue were administered to two hundred ten industrial school boys and twenty-five twelve year old children.

The Mentality of the Chinese and Japanese. Hsiao Hung Hsiao. The Journal of Applied Psychology, Feb., 1929, 9–31. A comparative report covering all psychological testing studies dealing with the Chinese and Japanese.

A Mental Alertness Examination for the Working Age Level. Richard S. Schultz. The Journal of Applied Psychology, Feb., 1929, 46–61. Two groups (three hundred ninety-two cases) ages ten to sixteen years were tested with Otis Intermediate, Viteles Mental Alertness Examination, and scholarship test for Grade VIIIB pupils. These correlations were secured: Mental Alertness Test (*a*) Reliability, .907; (*b*) validity against Otis criterion, .796; against school examination, .907; (*c*) correlation with grade, .575, with age, .144.

Measurement of Achievement

The College Psychology Test. J. E. Bathurst. The Journal of Applied Psychology, Feb., 1929, 81–85. A preliminary report also presenting a sample list of items.

Standardized Tests for Students of Education. Max D. Engelhart. Educational Administration and Supervision, Feb., 1929, 93–101. A review of available tests.

A Comparison of Objective Examination Methods. Harold Ellis Jones. Educational Method, Feb., 1929, 273–276. The author presents a number of the new types of tests, and reports the reliability for each type.

Some Significant Results of the Modern Foreign Language Study. V. A. C. Henmon. Journal of Educational Research, Feb., 1929, 79–91. The author presents graphs showing norms in French vocabulary, grammar, silent reading, and composition for high schools in United States, Canada and England. Wide variation in achievement between individual schools and classes is also noted; percentiles are used to show the varying emphases placed by different classes upon factors measured by tests.

The Correlation between the Difficulty of Latin Constructions and Their Frequency in High-school Latin. Mark E. Hutchinson. The Classical Journal, Mar., 1929, 412–420. With five hundred students, correlations between difficulty and frequency were for case constructions, .405 and .433; and for mood constructions, .104 and .174.

Psychology of Learning and School Subjects

Summary of Reading Investigations (July 1, 1927, to June 30, 1928). William Scott Gray. The Elementary School Journal, (1) Feb., 1929, 443–457.

Relative Effectiveness of Two Different Plans of Training in Silent Reading. Jesse D. Burks and Clarence R. Stone. The Elementary School Journal, Feb., 1929, 431–436. With four hundred four pupils in the fourth and fifth grade, no significant differences between groups (one trained by McCall "Standard Test Lessons in Reading," the other by Horn "Learn to Study Readers") were found when groups equated and finally tested on the Stanford Achievement Paragraph Meaning were also tested on the Chapman-Cook Speed of Reading Test.

Methods Employed to Stimulate Interests in Reading. III. William F. Rasche. The School Review, Mar., 1929, 204–214. A survey of methods.

The Nature of Difficulties Encountered in Reading Mathematics. J. S. Georges. The School Review, Mar., 1929, 217–226. An analysis of problems with which pupils had trouble revealed the following sources of difficulty: Difficulties in understanding and interpreting a statement (vocabulary and symbolism), thirty-seven per cent; inability to interpret mathematical processes and relationships, twenty-one per cent; etc.

Mistakes Which Pupils Make in Spelling. William F. Book and Richard S. Harter. Journal of Educational Research, Feb., 1929, 106–118. The writers analysed 5196 test papers, compositions and themes from Grades I to X, and from university freshmen. The analysis showed that errors were chiefly of these types: Phonetic spelling, omission of letter, repeating or adding a letter, ignorance of word, carelessness, and transposition of letters.

Scientific Pedagogical Foundation of Teaching Modern Foreign Languages. L. Silberberg. The Modern Language Journal, Feb., 1929, 376–392. A review of a number of scientific studies with a discussion of their application.

Problem Solving in Children and Adults. E. F. Heidbreder. Journal of Genetic Psychology, Dec., 1928, 522–544. Four groups (ten subjects each) were tested in ability to solve problems of varying difficulty—the problems consisted of locating an object concealed in one of a number of boxes which were presented in a certain order.

Character and Personality Traits

That Lenient Tendency in Rating. Natalie Kneeland. The Personnel Journal, Feb., 1929, 356–366. The means from ratings on a "Shopper's Scale," a "Rating Scale for Salespeople," and "Points to Be Considered in Judging a Department" were high, showing lack of discrimination by raters.

The Relative Reliability of Men and Women in Intuitive Judgments of Character. C. W. Valentine. The British Journal of Psychology, Jan., 1929, 16–238. A series of experiments in judging of traits shows little or no difference in this ability between men and women.

Statistical Methods

The Limits of Correlation between Three Variables. J. Ridley Thompson. The British Journal of Psychology, Jan., 1929, 239–252.

On the Sampling Theory of Intelligence. Stuart Carter Dodd. The British Journal of Psychology, Jan., 1929, 306–327.

The Pearson Formula, and a Further Note on the Kuhlman-Anderson Tests. F. Kuhlman. Journal of Applied Psychology, Feb., 1929, 32–45. The author concludes, on the basis of experimental evidence, that "no test can give uniformly good results at several different age levels" and that "the poorer the tests . . . the higher will be the Pearson r between the tests results and another criterion, and between the two halves of a battery of such tests."

Child Psychology

Child Development Abstracts and Bibliography. Committee on Child Development, National Research Council. Washington, Dec., 1928.

The Only Child. Norman Fenton. Journal of Genetic Psychology, Dec., 1928, 546–556. The study covered a group of one hundred ninety-three children (thirty-four only-children) from kindergarten to sixth grade who were rated by teachers in several traits; a second study concerned a group of five hundred twelve university students (seventy-three only-children) who were given the Woodworth Questionnaire. There were practically no differences between only-children and not-only-children.

Educational Psychology

Variability of Individual Performance. N. J. Lennes and Ira B. Fee. School and Society, Feb. 23, 1929, 264–268. A study of the variability of performance of one hundred thirty-five pupils of Grade VIII on arithmetical tests of the fundamentals.

The Effect of Overlearning on Retention. Journal of Experimental Psychology, Feb., 1929, 71–78. Over a one-day interval, increase in degree of learning from one hundred per cent to one hundred fifty per cent was followed by a corre-

sponding increase in retention; an increase from one hundred fifty per cent to two hundred per cent netted a decreasing increase in retention.

Primary, Recency, Frequency, and Vividness. Arthur Jersild. Journal of Experimental Psychology, Feb., 1929, 58–70. The writer tested 253 collegiate students, finding these methods to be most effective in oral presentation of a selection: five repetitions, score 244; four repetitions, 234; three repetitions, 211; primary first degree (statement placed first in selection), 192; "Now get this" (before a statement), 176; two distributed repetitions (eleventh statement and sixtieth statement in selection), 175; primary second degree, 163; two distributed repetitions (thirty-fifth and fortieth), 163; "Did you notice that?" (after statement), 158; pause, 143; recency (last statement in selection), 139; primary third degree, one hundred thirty-three; recency (second from last statement), 126; two concentrated repetitions (60–60), 126; loud statement, 123; bang on table, 121; recency (third from last statement), 121; gesture, 119; two concentrated repetitions (10–10), 119; normal, 104; slow, 82.

Controlled Experimentation in the Methods of College Teaching. Harl R. Douglass and others. University of Oregon Publication, Feb., 1929. Report of a series of experiments that indicated these conclusions: (*a*) The added hour a week devoted to quiz sections had no significant value. (*b*) Practically no difference in the effectiveness of the problem method of instruction as compared with the lecture method was found in a course in economics. (*c*) With students who had previously had psychology, the project method of study proved superior to the lecture method. (*d*) Independent study proved more profitable than the conventional method.

Concerning So-called Group Effects. Paul R. Farnsworth. Journal of Genetic Psychology, Dec., 1928, 587–594. With small groups of college students paired on Thorndike and Otis tests, no consistent effect in mean scores of working *with group* in comparison with working *alone* was found.

Curriculum

The Vocabulary Content of Compositions Written by College Students. Paul A. Witty and Mabel Fry. Journal of Educational Research, Feb., 1929, 135–138. Papers of three hundred forty freshmen, and junior and senior college students were collected. The students had been instructed to discuss the actual characteristics of their best-liked teachers, writing for twenty-five minutes. Little difference was found between upper classmen and freshmen, both groups exhibiting a paucity of vocabulary.

Objectives of Secondary Education According to the Opinions of Pupils. Lyle G. Thomson. The School Review, Mar., 1929, 198–302. A check list of objectives was submitted to three hundred forty-four pupils, Grades IX to XII. There was fairly close agreement between those objectives checked by pupils and those agreed upon by educational experts. "Citizenship" and "vocational" objectives were rated highest by pupils. There was no marked difference between sexes in traits ranked important.

Administration and Supervision

A Quantitative Investigation of the Efficiency of Instruction in High School Physics. Paul S. Miller. Journal of Educational Research, Feb., 1929, 119–127.

With the testing of six hundred students, the writer concluded that pupils have not acquired the facts; that the size of class has no consistent effect on score attained; that college preparatory curriculum group attained higher scores than normal, general, and mechanical curricula; that pupils who study physics in the fourth year in high school attain the highest scores; that teacher's marks and scores attained do not bear close relationship.

Reading Interests of Teachers. C. R. Roberts and Robert A. Davis. Educational Administration and Supervision, Feb., 1929, 102–116. An investigation (questionnaire) of the reading of two hundred forty elementary and secondary school teachers. Tables present distributions of time spent in professional, literature, current publications, and biography by groups classified according to teaching experience, subjects taught, age, and academic training.

An Objective Measure in Supervision. Doris F. Twitchell. Journal of Educational Research, Feb., 1929, 128–134. A form is presented, the desirable features being discussed in the article.

Participation in Testing Programs by Classroom Teachers. I. N. Madsen. Educational Administration and Supervision, Feb., 1929, 117–126. A study in errors in scoring of tests, and other features of the situation.

Administrative Practices in Dealing with Personnel Problems in Secondary Schools. Robert Woellner and W. C. Reavis. The School Review, Mar., 1929, 176–186. A study of practices in five hundred twenty-two secondary schools.

Records of Junior College Transfers in the University. Walter Crosby Eells. The School Review, Mar., 1929, 187–197. An analysis of the records of three hundred seventeen students from junior colleges showed this group to be superior to native Stanford student group in Thorndike test, in scholarship, in honors earned, and in entrance to graduate work.

Educational and Vocational Guidance

The Predictive Value of General Intelligence Tests in the Selection of Junior Accountants and Bookkeepers. Kenyon J. Scudder. The Journal of Applied Psychology, Feb., 1929., 1–8. With one hundred three cases replying to a questionnaire, the writer states that success in these occupations "depends to a large degree upon the mental capacity (Army Alpha score) of the individual."

Which Workers Have Good Attendance. C. J. Ho. The Personnel Journal, Feb., 1929, 385–389. A study of the association between lateness and absentism and these factors: age, marital status, sex, and personality.

Why This Career? Significance of Vocational Information in Decisions of College Students. R. B. Cunliffe. The Personnel Journal, Feb., 1929, 376–384. Factors influencing vocational choices of five hundred thirty-three students were studied by the questionnaire method. The most common factors mentioned were interest, conditions of work, income, school subjects, advice of parents or family, demand for workers.

A Technique for Job Analysis. Vera E. Dye and Edna W. Unger. The Personnel Journal, Feb., 1929, 367–375. Report of an attempt to place girls who were maladjusted and dull normal in suitable occupations.

The Long-time Effect of Freshman Counseling. P. W. Holaday. School and Society, Feb. 16, 1929, 234–236. The lowest decile (intelligence) of freshman

class was divided into a control group and an experimental group. The latter group was subjected to a program of coaching including questionnaires and interviews, remedial work on physical troubles, training in methods of study, improvement of reading speed, etc. In the experimental group there were fewer students dropped and higher grade points earned than in the control.

The Employment History of Auxiliary Pupils. Harold P. Thomas. The Vocational Guidance Magazine, Mar., 1929, 245–250. A report of one hundred forty-two pupils, regarding jobs held, periods of work and idleness, sex differences and kinds of employment.

Miscellaneous

The Effect of the Quality of Penmanship on Grades. Journal of Educational Research, Feb., 1929, 102–105. With four hundred fifty teachers as graders, the average grade as a composition of the good penmanship paper (90, Ayers Scale) was ten points higher than that of the poor penmanship paper (30, Ayers) of identical content.

Women Taxicab Drivers. Sex Differences in Proneness to Motor Vehicle Accidents. Morris S. Viteles and Helen M. Gardner. The Personnel Journal, Feb., 1929, 349–355. Although other investigations tend to show greater accident rates for men than for women, this study shows a higher accident rate for women cab drivers than for men drivers.

A Study and Investigation of the Health of Women Teachers of Physical Education. Kathro Kidwell and Dorothy Simpson. American Physical Education Review, Feb., 1929. Report of questionnaires submitted to four hundred graduates in physical education.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

Educational Yearbook of the International Institute of Teachers College, Columbia University, 1927, by I. L. Kandel, (Ed.) New York; Teachers College Bureau of Publications, 1928. Pp. XV + 618.

Each year, under the editorship of Dr. Kandel, the International Institute of Teachers College surveys the educational systems of about a score of countries extending from China to Peru. Some of the more important ones, like those of the United States, England, France and Germany, get surveyed each year; others have to take their turns. Up to the present about fifty systems have been described, the whole constituting a veritable gold mine of information for all who are interested in comparative education. This year the countries dealt with include Belgium, Denmark, England, France, Germany, Hungary, Iraq, Lithuania, Mexico, Russia, Serbia, Croatia and Slovenia, Spain, Turkey and the United States. The contributors, so far as they are known to the reviewer, are the best each of the countries affords. In addition, Part II of this year's volume deals with the almost insoluble problem of teacher-training, with important contributions to the subject by T. Percy Nunn, Félix Pécaut, H. Theodore Becker, and Wm. C. Bagley.

Each contributor, it seems, is asked by the editor to report on the following educational features of his country—administration of the system, pre-school and kindergarten education, elementary, secondary and higher education, commercial and vocational education, preparation of teachers, educational finance and special features such as the use of tests and scales. This scheme gives a unity to the various reports, although variety of necessity creeps in and gives a welcome tang to the whole.

The editing, on the whole, is so well done that it almost appears captious to point out the places where improvements could still be made. Yet the reviewer finds so much difficulty with the various monetary systems that he feels sure they ought to be brought to the

common dollar standard. How many readers can translate roubles, kronen, pounds, pengos, pesetas, pesos, rupees, francs (Belgian and French) and lites as he goes along? What is a lite anyway? Even in those cases where a translation has been made (pounds to dollars) it has been incorrectly done. The bibliographies are somewhat unequal and more attention should be devoted to them.

The reviewer enjoyed most the accounts of the Iraq, Lithuanian, Mexico and Russian systems, and the whole of Part II, which is most illuminating, but other readers might pick out a very different set for special commendation.

As in previous years the reviewer is mute with astonishment at the linguistic abilities of an editor who translates articles from the French, German, Lithuanian, Spanish and Jugo-Slavian tongues. The volume well deserves its place in the list of the best sixty educational works published during the year. PETER SANDIFORD.

University of Toronto.

Overlappings and Omissions in Our Courses of Study, by W. J. Osburn. Bloomington, Ill.: Public School Publishing, Co., 1928.

The placement of units in the various grades is an important curriculum problem. Mr. Osburn's study does not aim to solve the problem but points out specifically the overlapping of elementary, secondary, and higher education in English composition, United States History, Physics and General Science. Overlapping is limited to those facts which were presented two or more times in the same relations or sets. The unit of measurement used was the question and the material study was the textbook. Each sentence in the sources on the three levels was converted into a question. The questions were then compared to discover duplication. English composition was studied in three states; United States history in four States; and physics and general science in ten states.

Mr. Osburn estimates that the amount of overlapping is equal to a whole year in about six. A large amount of the duplication is factual in character. There is a complete absence of planning in the duplication which occurs. The study includes tables giving the amount of overlapping. The major part of the monograph is devoted to a detailed enumeration of a portion of the facts which are duplicated in each subject.

From the point of view of modern curriculum making, this study emphasizes the need of solving three important problems: (1) To determine whether the period of formal education may be reduced; (2) to effect a closer relationship between elementary, secondary and higher education; (3) to provide for planful repetition of essentials, which, negatively stated, means to eliminate wasteful duplication.

The author, in my opinion, undertook too big a task. Consequently certain portions of the study are fragmentary. Many elaborate techniques are introduced in insignificant portions of the study that take the reader off into involved by-paths. The author has a tendency to be garrulous which irritates the busy student of education. The monograph on the whole is a carefully prepared study revealing the desultory selection and placement of units of the curriculum on the elementary, secondary, and higher levels of education.

HENRY HARAP.

School of Education, Western Reserve University
Cleveland, Ohio.

Intelligence Tests. Their Significance for School and Society, by Walter F. Dearborn. Boston: Houghton Mifflin Co., 1928. Pp. XXIV + 326.

Many volumes have been published on the subject of intelligence tests, their characteristics, uses and merits. Too often the view point of the author has been non-critical, and too frequently highly speculative conclusions have been reached.

Professor Dearborn's book, consisting largely of material first presented as a series of lectures delivered by the author at the Lowell Institute, Boston, in 1925, is a critical evaluation of intelligence tests. The author appreciates the advantages to be derived from intelligence tests, but as the result of extensive study, he is aware of limitations in their use and believes that many fallacies in interpretation have resulted.

The book includes, in addition to an appraisal and thoroughgoing criticism of test interpretation, a description of intelligence tests, a discussion of the nature of intelligence, and a brief history of early attempts to measure intelligence. Among the topics discussed are: The nature of intelligence and the influence of schooling, environment, maturity, upon it; the relation of intelligence to school achievement; special abilities and abilities considered in the light of intelligence

tests, provisions in the schools for pupils of differing ability; the relation of intelligence and schooling to behavior; social and educational policies in the light of intelligence tests.

Professor Dearborn maintains that there are other sorts of intelligence not adequately measured by intelligence tests. He cites evidence to show that intelligence can be increased by favorable environmental conditions. Other fallacies pointed out are chiefly of one type—cause and effect relations deduced from data that are merely associated, as for example, the relation of such factors as quality or extent of schooling to intelligence, as measured by tests. In some cases the author does no more than open the question for discussion; in other cases he presents evidence to disprove a conclusion or to prove the validity of an opposite conclusion. For the most part he refrains from dogmatic assertions as to the truth or falsity of conclusions drawn from the use of intelligence tests. Occasionally, however, he indulges in far-fetched arguments himself, without sufficient evidence to support his position.

In his discussion of word-blindness and reading, the author has emphasized mechanical features of doubtful significance in the reading process, and has apparently disregarded the more significant factors of instructional method, the child's concept of the reading process, and his preparation and readiness for reading activities, as probable factors in causing his difficulty. There is insufficient evidence concerning the relationship of handedness, reading, writing, and speech. Improvement in reading is not proved as the reason for an increase in IQ from one hundred four to one hundred twenty-six in the case of one child described. Many cases are on record of even greater gains in IQ accompanied by no change whatsoever in reading status.

Many of the fallacies pointed out by the author do not represent the best thinking of leading psychologists, test users and makers. The most competent and scholarly psychologists have seldom subscribed to conclusions resulting in the more extremely fallacious arguments to which the author calls attention.

On the whole, the material presented is worthwhile. It will prove valuable, if as a result, more caution on the part of test users is exercised in subscribing to generalizations from intelligence test data.

GERTRUDE H. HILDRETH.

Lincoln School of Teachers College.

Fundamental Principles of Learning and Study, (Revised Edition), by A. S. Edwards. Baltimore; Warwick and York, Inc., 1925. Pp. 255.

This is a revision of the writer's work published in 1920. The organization is considerably changed and improved, and the whole book is set up in a new style of type.

Its chief characteristics are: (1) It emphasizes applications of psychology to study and learning; (2) two chapters of concrete cases of behavior highly suggestive for teachers and students are included; (3) one chapter contains a multitude of brief rules for study for the use of the supervised study teacher; (4) some readers may criticize the possible overemphasis of the *habit* concept; (5) as a text for how-to-study courses it is psychological only, and does not attempt to tell how to do specific school tasks. In other words, it is intended primarily as a psychology text, emphasizing study methods.

C. C. Crawford.

University of Southern California.